

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5416.83 Seconds  
(without alignments)  
17274.699 Million cell updates/sec

Title: US-09-728-552a-3\_COPY\_1\_2000

Perfect score: 2000  
Sequence: 1 gatctctccgcctcagcct.....ctcgaataatagatgtgtg 2000

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hnc:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415.8	20.8	432	9	AQ015472 CIT-HSP-2
2	413	20.6	8158	4	CR749268 Homo sapi
3	399.2	20.0	5937	4	BSM807294 Homo sapi
4	385	19.2	707	5	BU616791 UI-H-FH1-
5	373	18.6	700	5	BU622317 UI-H-FH1-
6	360.8	18.0	690	6	CA430123 UI-H-FH1-
7	360.4	18.0	9266	6	HSB806813 Homo sapi
8	351.6	17.6	994	6	BSG13167 AGENCOURT
9	351.6	17.1	814	1	AU120614 UI-H-FH1-
10	341.2	17.1	814	1	AU120614 UI-H-FH1-
11	335.6	16.8	719	5	BSX55959 DKEP2781A
12	335.2	16.8	907	6	CF242966 AGENCOURT
13	330.6	16.5	603	9	AQ320818 RPII1-92
14	330	16.5	687	5	BZ603609 WHA4V1TF
15	330	16.5	718	5	BSX495939 DKEP2781A
16	329.8	16.5	721	7	CR774535 Pongo pyg
17	328.2	16.4	831	7	CR859436 Pongo pyg
18	327.4	16.4	5336	4	BSG476842 602525095
19	326.6	16.3	929	2	CE040574 tigr-gss-
20	325.6	16.3	755	9	AQ237450 RPII1-61
21	325.6	16.3	815	9	BU622127 UI-H-FH1-
22	325.4	16.3	705	5	BU622127 UI-H-FH1-
23	325.2	16.3			AQ748474 HS-5540 A
24	325	16.2			CC478317 CH240_304
25	324.2	16.2			B0067736 AGENCOURT
26	321.4	16.1			AQ750123 HS-5575 A
27	318	15.9			BG028427 602294332
28	316.4	15.8			CR851333 UI-CF-FNO
29	315.8	15.8			CF126937 UI-HF-ETO
30	314.8	15.7			CG744253 RPI1-5711
31	312.6	15.6			BU740723 UI-E-BUO-
32	311.8	15.6			BU740723 UI-E-BUO-
33	311.8	15.6			BU740723 UI-E-BUO-
34	311.6	15.6			BU740723 UI-E-BUO-
35	311.4	15.6			BU740723 UI-E-BUO-
36	311.2	15.6			BU740723 UI-E-BUO-
37	310.6	15.5			BU740723 UI-E-BUO-
38	310.4	15.5			BU740723 UI-E-BUO-
39	306.6	15.3			BU740723 UI-E-BUO-
40	306	15.3			BU740723 UI-E-BUO-
41	305.6	15.3			BU740723 UI-E-BUO-
42	305.4	15.3			BU740723 UI-E-BUO-
43	304.2	15.2			BU740723 UI-E-BUO-
44	304	15.2			BU740723 UI-E-BUO-
45	303	15.2			BU740723 UI-E-BUO-
46	301.2	15.1			BU740723 UI-E-BUO-
47	299	14.9			BU740723 UI-E-BUO-
48	298.8	14.9			BU740723 UI-E-BUO-
49	297.8	14.9			BU740723 UI-E-BUO-
50	297.4	14.9			BU740723 UI-E-BUO-
51	295.8	14.8			BU740723 UI-E-BUO-
52	295.4	14.8			BU740723 UI-E-BUO-
53	293.8	14.7			BU740723 UI-E-BUO-
54	293.4	14.7			BU740723 UI-E-BUO-
55	293.2	14.7			BU740723 UI-E-BUO-
56	292.4	14.6			BU740723 UI-E-BUO-
57	290.4	14.5			BU740723 UI-E-BUO-
58	289.8	14.5			BU740723 UI-E-BUO-
59	289.8	14.5			BU740723 UI-E-BUO-
60	289.8	14.5			BU740723 UI-E-BUO-
61	289	14.4			BU740723 UI-E-BUO-
62	289	14.4			BU740723 UI-E-BUO-
63	289	14.4			BU740723 UI-E-BUO-
64	288.8	14.4			BU740723 UI-E-BUO-
65	286.8	14.3			BU740723 UI-E-BUO-
66	285.8	14.3			BU740723 UI-E-BUO-
67	284.6	14.2			BU740723 UI-E-BUO-
68	284.2	14.2			BU740723 UI-E-BUO-
69	283.4	14.2			BU740723 UI-E-BUO-
70	282.4	14.1			BU740723 UI-E-BUO-
71	281.6	14.1			BU740723 UI-E-BUO-
72	281.4	14.1			BU740723 UI-E-BUO-
73	280.4	14.1			BU740723 UI-E-BUO-
74	280.4	14.0			BU740723 UI-E-BUO-
75	279.8	14.0			BU740723 UI-E-BUO-
76	278.6	13.9			BU740723 UI-E-BUO-
77	278.4	13.9			BU740723 UI-E-BUO-
78	278.4	13.9			BU740723 UI-E-BUO-
79	278.2	13.9			BU740723 UI-E-BUO-
80	278.2	13.9			BU740723 UI-E-BUO-
81	278.2	13.9			BU740723 UI-E-BUO-
82	277.2	13.9			BU740723 UI-E-BUO-
83	277.2	13.9			BU740723 UI-E-BUO-
84	276.6	13.8			BU740723 UI-E-BUO-
85	275.8	13.8			BU740723 UI-E-BUO-
86	275.4	13.8			BU740723 UI-E-BUO-
87	274.6	13.7			BU740723 UI-E-BUO-
88	274.6	13.7			BU740723 UI-E-BUO-
89	273.8	13.7			BU740723 UI-E-BUO-
90	272.6	13.6			BU740723 UI-E-BUO-
91	272.2	13.6			BU740723 UI-E-BUO-
92	272	13.6			BU740723 UI-E-BUO-
93	271.8	13.6			BU740723 UI-E-BUO-
94	271.6	13.6			BU740723 UI-E-BUO-
95	271.4	13.6			BU740723 UI-E-BUO-
96	271.4	13.6			BU740723 UI-E-BUO-
97	271.4	13.6			BU740723 UI-E-BUO-
98	271.4	13.6			BU740723 UI-E-BUO-
99	271.4	13.6			BU740723 UI-E-BUO-
100	271.4	13.6			BU740723 UI-E-BUO-





## ORIGIN

LFIVHORIHTGEKPYOCEKCAFTOKIASIOHQRVHTGEKPYEKCVCAGAFKMGSE  
VOHOKLHPVEKPKVPLVSGPQCSFPAIPVLLQSGSCASAVAPSLTFPHAVLI  
PTSGNFFMLPTSGIPSSAQIVRVFGLPVTKPSVLLTPSSHS"

Query Match 20.6%; Score 413; DB 4; Length 8158;  
Best Local Similarity 73.2%; Pred. No. 1.5e-55;  
Matches 686; Conservative 0; Mismatches 195; Indels 56; Gaps 10;

1062 GTGAGTCACACAGTCTTTGGCTCCAGTCATATATAAGTTTGGCTTATCTACT 1121  
1105 GGAAGTCACACATTTATTTATTTTCCAGTACCTATACAGTATATCTATAT 8046  
1122 GTAGTCTGTAAGTGCATAGTATGCTTAAATAAACAATACCTTAAATTTTAAA 1181  
8045 ATAGTCTATTAAGTTGCATATGATATATCTGACATGATACCTTAAATTTTAAA 7986  
1182 TGCTTATTAATTAATAAATGCTTAAATGATGATGATGATGATGATGATGAT 1241  
7985 TATTTATTTGCTAAAGATGCTTAAATGATGATGATGATGATGATGATGAT 7931  
1242 GCTGTCGGAAGTCTTTTCTTATGATGATGATGATGATGATGATGATGAT 1290  
7930 GCTGGAAGACAGTCTTGCCTCAGCTGATGATGATGATGATGATGATGATGAT 7871  
1291 GAAGCTTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1345  
7870 GAAGCTTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7811  
1346 AGTTCATCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1405  
7810 AACTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7752  
1406 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1464  
7751 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7692  
1465 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1524  
7691 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7632  
1525 GATTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1584  
7631 GATTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7576  
1585 AAGAACATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1644  
7575 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7516  
1645 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1704  
7515 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7461  
1705 CCTTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1764  
7460 ACTTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7401  
1765 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1821  
7400 TATTCACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7341  
1822 ATGTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1881  
7340 ATGTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7281  
1882 TCCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1941  
7280 CC-----ACTATGCGACCCACAGCTTACAAATATATATCTTAAAT 7239  
1942 AATAGCTTGAAGTGAATTAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1978  
7238 AATA---CTTGAAGTCAAAATTAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7205

## RESULT 3

HSM807294

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

CONSRTM

TITLE

JOURNAL

COMMENT

FEATURES

source

HSM807294 5937 bp RNA linear HTC 20-JAN-2005

Homo sapiens mRNA; cDNA DKFZp686B15223 (from clone DKFZp686B15223).

BX647150.1 GI:34366178

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 5937)

Wambutt, R., Heubner, D., Mewes, H. W., Weil, B., Amid, C., Osaenger, A.,

Fodor, G., Han, W., and Wiemann, S.

The German cDNA Consortium

Submitted (20-JAN-2005) MIPS, Ingolstaedter Landstr. 1, D-85764

Neuberberg, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;

sequenced by Agowa (Berlin/Germany) within the cDNA sequencing

consortium of the German Genome Project.

This clone (DKFZp686B15223) is available at the RZPD Deutsches

Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.

Please contact RZPD for ordering:  
http://www.rzpd.de/cgi-bin/products/c1.cgi?cloneID=DKFZp686B15223

Further information about the clone and the sequencing project is

available at <http://mips.gsf.de/projects/cdna/>.

Location/Qualifiers

1..5937

/organism="Homo sapiens"

/mol\_type="pre-RNA"

/db\_xref="RZPD:DKFZp686B15223Q"

/db\_xref="taxon:9606"

/clone="DKFZp686B15223"

/cistron\_type="small intestine"

/clone\_lib="686 (synonym: hicc3). Vector pSport1\_Sfi; host

DH10B; sites SfiIa + SfiIb"

/dev\_stage="adult"

/note="CD8 beta-chain glycoprotein, N-terminus truncated,

not fully spliced"

## ORIGIN

Query Match 20.0%; Score 399.2; DB 4; Length 5937;  
Best Local Similarity 70.7%; Pred. No. 2.4e-53;  
Matches 728; Conservative 0; Mismatches 233; Indels 69; Gaps 12;

975 AGTACGCGATACCTTGGAGATACGTGGGTTGGTTCCATACACCAATAATACAA 1034  
581 AGTATGGGAAACCTTGAAGATATCTCAGGTTCAAGTTCCAGACCAACCAATAACCA 639  
1035 TATGCAAGAAGTGATATACATAAAGTGAAGTCAACAAGTCTTTGGTCCAGTGC 1094  
640 -----ATATCAATTAAGCCAGTTAATTCATTTTGGTTCCAGTGT 685  
1095 ATATTAAGTTTGGTTATATACATCTGATGCTGTTAAGTGCATAGTATATGCT 1154  
686 ACATTAAGTTATGTTTACCTATAGTATAGTCTACTAAGTGCACAAAGCATTAATGCT 745  
1155 AAAAAAACAATAC--CTTAATTTTAAATGCTTTATTAATAAATAAGTAAACATCAT 1212  
746 GAAAAAATAAGTACGCTTATTAATAAATAATTTTAAATTAATCATGCTTAAATAC 805  
1213 TTGAGCATTCAGAGTGTGAATCTTTTGGTGGGAGAGTCTTT-----C 1260  
806 CTGAGCTTCAGTGAATCATTAATTTCTTGCTGATGAGGAGCATCTTGCTTGATGCA 865  
1261 TTATGATGATGATGATG--GGGGGTCAAGTCTGAAGTGAAGTGGTGGAGTTTCTT 1319  
866 TGGCTGCTACATGATCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 925

```

QY 1320 AAA-----ACACAGGAGGAGATTGCAATATTCAGTTGACTCTTCTCTTCAAGAAATTC 1374
Db 926 AAAATAGGCAACAAACAAATCTTGGCACATTGATTGACTCTTCTTCAAAAAGTTTA 985
QY 1375 TCTCTAGTGTGATGCTTTTGTGATGAGATTTTATGACAGTAGAACTTCTTGAAT 1434
Db 986 TCTGTAGATGAGATGCTTTGATGATGATTTTACTACAGTAGAAACAGATTTCAAAAT 1045
QY 1435 GGA-TCATCTCTCAAAACCTGCTGCTTTTAAACAACCTTAAGTTATTAATTCGA 1493
Db 1046 GGAGTCACCTCTCTCAAAATACGAGCCCTCTTCTCACTAAGTTATGTAACATTTCT 1105
QY 1494 ATCCATGTTGATTCATTCAAAATTTTCAAGTGTCTTACCGAGAGATTCATCTC 1553
Db 1106 ATCTTCTGTGTATTTCAACATTTTCAACATCTTCAACGAGAGTATTCATTTTC 1165
QY 1554 ATTTCTGAGATGGAATCTTTGCTCATTCATTAAGAAATTCCTCATCTTTCAAGTT 1613
Db 1166 AAGAAACCACT-----TCTTGTCTATCAGAGAGAAATTAACCTCATGTGT -AAGTTT 1220
QY 1614 TATCATGAGATTCAGACAAATACAGTATGTTCTTACGCTCATCTTCACTTTAATTCAG 1673
Db 1221 GATCAGGGAATGCAACAAATTCAGCAAGCTTTCAG-----ACTCCACTTATCTCTAG 1275
QY 1674 TTCTCTGCTGTTCTTACCAACATCTGAGTTCCTCTCATTTGAAGTCTTGAACCTTC 1733
Db 1276 TTCTCTGCTGATTTTCAACATCTTCAAGTATCTTCTCTCACTGAAGTCTTGAACCTTC 1335
QY 1734 CAAGTCATTCATGAGGATGGAATTCGATCTTTCCTCAAAATTCCTGTTAATTTAATTT 1793
Db 1336 CAACTCATTCAGAGAAATGGAATTCAGTCTTCCAAACCTCTCTAATGTTGATTTA 1395
QY 1794 GAC---CTCCATGATCATGAATGTTCTTAAATGAGCACTGAGATGTAATCTTTCCA 1850
Db 1396 TACCTCTCTCCATGAATTCACAAATGTTCTTGTGATTCAGATGTTGAATCTTTCCA 1455
QY 1851 AAGGTTTCAATTTCTTGTAGTCAGATTCATCATCAGAGATTCATCTTCAATGCCA 1910
Db 1456 GAAGGTTTCAACCAAGATCTA-----TCAAGGATATATATCTGTGGCA 1499
QY 1911 GTTATAGCCTTATGATGATTTTCTTCAATTAATAGGCTTGAAGTTGAATTAATCTCT 1970
Db 1500 GCTTACCTTATGAGAAATGATTTCT-----TAAATGAATTTGAAGTCAAAATTAATCTCT 1555
QY 1971 TGATTCATTT 1980
Db 1556 TGGTTCATGT 1565

```

```

RESULT 4
BUE16791 707 bp mRNA linear EST 23-SEP-2002
LOCUS UI-H-FH1-bfj-k-05-0-UI.s1 NCI CGAP FH1 Homo sapiens cDNA clone
DEFINITION UI-H-FH1-bfj-k-05-0-UI 3', mRNA sequence.
ACCESSION BUE16791.1 GI:23283006
VERSION BUE16791.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE 1 (bases 1 to 707)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
        Tumor Gene Index
        Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
        Tissue Procurement: James Martin
        cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
        cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
        DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

```

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@iowa.edu  
The following repetitive elements were found in this cDNA sequence: 20-707, >ITGGER1#DNA/MER2\_type (matched complement)  
Seq primer: M13 FORWARD  
POLYA=yes.

#### FEATURES

```

source
1..707
location/Qualifiers
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-H-FH1-bfj-k-05-0-UI"
    /tissue_type="Cell Line"
    /dev_stage="Adult"
    /lab_host="DH10B (Life Technologies)"
    /clone_lib="NCI CGAP FH1"
    /note="Organ: Chondrosarcoma; Vector: pT73-Pac (pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AAGATCCGGC. The cell line was provided by Dr. James Martin from the University of Iowa.
    TAG TISSUE=Human Chondrosarcoma Cell line C58 - Grade 1
    Chondrosarcoma
    TAG_LIB=UI-H-FH1
    TAG_SEQ=AAGATCCGGC"

```

#### ORIGIN

```

Query Match 19.2%; Score 385; DB 5; Length 707;
Best Local Similarity 79.3%; Pred. No. 6,4e-51;
Matches 554; Conservative 0; Mismatches 115; Indels 30; Gaps 7;
QY 1170 TTAATTTTAAATGCTTTTATTAATAAAATGCTAAATC-ATTGAGCATTCAGTAG 1228
Db 18 TAAATTTTAAATTAATTTATGCTTAAATAAATGCTGATTCATTCAGAGCTTCGGTAG 77
QY 1229 TTGTAATCTTTTTCGTCGTCGTCGTCGTCCTTCTTATGATGATGAT-----CG 1277
Db 78 TCGTACTCTTTTTCGTCGTCGTCGTCGTCCTGTCGAGTGTGCTGACTGTCAG 137
QY 1278 GGGGTACAGTGTGAGGCTTAGGGTGTGTCGTCAGTTCCTTAA-----ACAAAGTGA 1332
Db 138 GGTGTGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 197
QY 1333 AGATTCGAATATGATGATGATCTTCTTCATGAAAGATTTCTCTAGTGTGATGCT 1392
Db 198 AGCTTTTCGATAGATGATGATCTTCTTCATGAAAGATTTCTCTAGTGTGATGCT 257
QY 1393 TTTTATGATGATTTTATGACAGTAGAACTTCTTGAATTTGA-TCATCTCTCAAA 1451
Db 258 GTTTGATGATGATTTTGGCCACAGTAGAGCTTCTTCAAAATTTGAGTCAATCTCTCAA 317
QY 1452 CCGTGTCTGCTTTTAAACCTAAGTATATATATTTCTGAATCCATGTCATTTCTTTC 1511
Db 318 CCGTGTCTGCTTTTAAACCTAAGTATATATATTTCTGAATCCATGTCATTTCTTTC 377
QY 1512 AACATTTTCAAGTGTCTTACACAGAGTAGATTCATCTCATTTCTTCTGATGATC 1571
Db 378 AACATTTTCAAGGATCTTCAACAGGATGATTCATCTCAAGAAACCACTT-----TC 433
QY 1572 TTGCTCATTCATTAAGAAATTTCTCATCTGTTCAAGTTTATCATGATGATGACGA 1631
Db 434 TTTCGATCATTCATTAAGAAACCACTTCATCTGTTTAAAGTTTATCATGATGATGACGA 493

```

QY 1632 ATACAGTCATGCTTCAGGCTTCCTTCATTTTAATTCAGTCTCTGCTGTTTAC 1691  
 DB 494 ATTCAGTTACATCTTCAAG-----CTCTACTCTTAATTCAGTCTTCTGTTATTCAC 548  
 QY 1692 CACATCTGTGCTTCTCTTCATTTGAAGTCTTGAACCTCCAGATCATCATGAGGT 1751  
 DB 549 CACATCTGCATCTTCTCTCCTCAGAGTCTTGAACCCCTCATATGATCATCAATGAGGT 608  
 QY 1752 TGAATGAGCTTCTCAAAATTCCTGTAATATATTAATTTGA---CTCCCATGATC 1808  
 DB 609 GGGATCACTTCTTCCAACTCTGTAATGTTGACATTTGACCTCTCCCATGATC 668  
 QY 1809 ATGAATGTTCTTAATGACACCTGGAATGATCTTT 1847  
 DB 669 ACGAATGTTTTTAATGCGATCTAAATGATGATCTTT 707  
 RESULT 5  
 BU622317 700 bp mRNA linear EST 23-SEP-2002  
 LOCUS BU622317  
 DEFINITION UI-H-FH1-bfu-h-08-0-UI.s1 NCI CGAP\_FH1 Homo sapiens cDNA clone  
 ACCESSION BU622317 GI:23288532  
 VERSION BU622317.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 700)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: James Martin  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@iowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 20-695, >RIGER1#DNA/MER2\_type (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=yes  
 location/Qualifiers  
 1..700  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FH1-bfu-h-08-0-UI"  
 /tissue\_type="Cell Line"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI CGAP FH1"  
 /note="Organ: Chondrosarcoma; Vector: pT73-Pac  
 (Pharmacia) with a modified polylinker; Site 1: EcoR I;  
 Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library  
 obtained from a cell line derived from grade I  
 chondrosarcoma tissue. The library was constructed and  
 normalized according to Bonaldo, Lennon and Soares, Genome  
 Research, 6:791-806, 1996. First strand cDNA synthesis was  
 primed with an oligo-dT primer containing a Not I site.  
 Double stranded cDNA was ligated to an EcoR I adaptor,  
 digested with Not I, and cloned directionally into  
 pT73-Pac vector. The oligonucleotide used to prime the  
 synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 AGAATCCGCG. The cell line was provided by Dr. James Martin

from the University of Iowa.  
 TAG TISSUE=Human Chondrosarcoma Cell Line C88 - Grade 1  
 Chondrosarcoma  
 TAG LIB=UI-H-FH1  
 TAG\_SEQ=AGAATCCGCG"

ORIGIN  
 Query Match 18.6%; Score 373; DB 5; Length 700;  
 Best Local Similarity 78.9%; Pred. No. 5e-49;  
 Matches 542; Conservative 0; Mismatches 115; Indels 30; Gaps 7;

QY 1170 TTAATTTAAAGCTTTTAACTAAATAAAATGCTAACATC-ATTGAGCATTCAGTGA 1228  
 DB 18 TAAATTTAAATAATCTTATTTGCTAAATAATGCTAATTAATCAATCAGCCTTGAGAG 77  
 QY 1229 TTGTAATCTTTTGTCTGTGGAAGTCTTTCTTATTAATGATGATGAT------CG 1277  
 DB 78 TCGTACTCTTTTGTCTGTGGAAGGCTCTTCTCGAGGTTGATGTTGCTGACTGCTCAG 137  
 QY 1278 GGGGTGAGTGTGGAAGCTTAAGGCTGTGAGGAGTTGCTTAA-----ACAACAGTGA 1332  
 DB 138 GGTGTGTGCTGTGAGAGTTGGGTGTGCTGTGCAATTTCTTAAATTAAGACATCGGTGA 197  
 QY 1333 AGATTTGCAATATGATGATGATCTTCTTCATGAAAGATTTCTCTAGTGTGATGCT 1392  
 DB 198 AGCTTTTGCATGATGATGATCTTCTTCATGAAAGATTTCTCTAGTGTGATGCT 257  
 QY 1393 TTTTATGATGATTTTATGACACAGTGAACCTTTTGAATAATGGA-TCAATCTCTCAA 1451  
 DB 258 GTTGTATGATGATTTTGTGCTGCAAGTGAAGCTTCTTCAAAATGGAATCAATCTCTCAA 317  
 QY 1452 CCCGTGCTGCTTTTAAACCTAAGTTAATATAATTCGATTCATTCGATTC 1511  
 DB 318 CCGTGTGCTGCTTTTAACTAAATTAATTAATTAATTCGATTCATTC 377  
 QY 1512 AACAATTTTCAAGTGTCTTCAACAGAGTGAATTCATCTCATTTCTCGAGATGAAATC 1571  
 DB 378 AACATGTTTCAAGCATCTTCAACAGAGTGAATTCATCTCATTTCTCGAGATGAAATC 433  
 QY 1572 TTTGCTCATTCATTAAGAAATTTCTCATCTGTTCAAGTTTATCATGATTCAGCA 1631  
 DB 434 TTGGCGCATCATTAAGAAACATCTCCATCTGTTAAAGTTTATCATGATTCAGCA 493  
 QY 1632 ATACAGTCATGCTTCAGGCTCAGTTCATCTTTAATTCAGTCTCTGCTTCTTAC 1691  
 DB 494 ATTCAGTTACATCTTCAAG-----CTCTACTCTTAATTCAGTCTTCTGTTATTCAC 548  
 QY 1692 CACATCTGTGCTTCTCTTCATTTGAAGTCTTGAACCTTCGAAGTCAATGAGGT 1751  
 DB 549 CACATCTGCATCTTCTCTCCTCAGAGTCTTGAACCCCTCATATGATCATCAATGAGGT 608  
 QY 1752 TGAATGAGCTTCTTCCAAATTCCTGTAATATTAATTTTGAATCTC--CATGATC 1808  
 DB 609 GGGATCACTTCTTCCAACTCTGTAATGTTGACAAATTTGACCTCTCCATGATC 668  
 QY 1809 ATGAATGTTCTTAATGACACCTGGAAT 1835  
 DB 669 ACGAATGTTTTTAATGCGATCTAAAT 695  
 RESULT 6  
 CA430123 690 bp mRNA linear EST 07-NOV-2002  
 LOCUS CA430123  
 DEFINITION UI-H-FH1-bfu-h-14-0-UI.s1 NCI CGAP\_FH1 Homo sapiens cDNA clone  
 ACCESSION CA430123  
 VERSION CA430123.1 GI:24792849  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1. (bases 1 to 690)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: James Martin  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
The following repetitive elements were found in this cDNA  
sequence: 20-690, >TIGER1#DNA/MER2\_type (matched complement)  
Seq primer: M3 FORWARD  
POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
1. 690  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-FH1-bis-f-14-0-UI"  
/tissue\_type="Cell Line"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP FH1"  
/note="Organ: Chondrosarcoma; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: BcoR I; Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCCGCGC. The cell line was provided by Dr. James Martin from the University of Iowa.  
TAG TISSUE=Human Chondrosarcoma Cell Line C8 - Grade 1  
Chondrosarcoma  
TAG\_LIB=UI-H-FH1  
TAG\_SEQ=AGAATCCGCGC"

## ORIGIN

Query Match 18.0%; Score 360.8; DB 6; Length 690;  
Best Local Similarity 79.0%; Pred. No. 4.2e-47;  
Matches 539; Conservative 0; Mismatches 112; Indels 31; Gaps 8;

1170 TTAAATTTAAATGCTTTTAAATGCTTAACATC-ATTGAGCATTCAGTGAG 1228  
18 TAAATTTAAATGCTTTTAAATGCTTAACATC-ATTGAGCATTCAGTGAG 77  
1229 TTGTAATCTTTTCTGCTGAGAGGCTTTTCTTATGAGACTGAT-----CG 1277  
78 TCGACTCTTTTCTGCTGAGAGGCTTTTCTTATGAGACTGATGCTGAG 137  
1278 GGGGTGAGGCTGAGAGGCTGAGAGGCTTCTTAA-----ACACAGAGGA 1332  
138 GGTGGGTGAGGCTGAGAGGCTGAGAGGCTTCTTAAATGAGCATTCGAGTA 197  
1333 AGATTGCAATATCAGTGAATCTTCTTTTCTGAGAAAGATTTCTCTAGTGTGATGCT 1392  
198 AGCTTTTCGATGAGTGAATCTTCTTTTCTGAGAAAGATTTCTCTAGTGTGATGCT 257  
1393 TTTTGATGAGATTTTATGACAGTGAAGATCTTTTGAATTTGA-TCATCTCTCTCAA 1451  
258 GTTGTATGAGATTTTTCACAGTGAAGATCTTTTCAAAATTTGAGTCAATCTCTCAA 317

QY 1452 CCGTCTGCTTTTAAACCACTTAAGTAAATATATTTCTGAATCATTTGTCTATTC 1511  
D 318 CCTGCTGCTGCTTTTAACTAAGTAAATATATTTCTTAATCTTTGTGTCTATTC 377  
QY 1512 AACATTTTACAGAGTCTTCAACGAGTGAATTCATCTCATTTCCGAGATGGAATC 1571  
D 378 AACATTTTACAGAGTCTTCAACGAGTGAATTCATCTCATTTCCGAGATGGAATC 433  
QY 1572 TTTGCTCATCTTAAGAAAGAAATTCCTCATCTGTTCAAGTTTATCATGATTCAGCA 1631  
D 434 TTTGGCATCTCAATGAAACCACTCTCATCTGTTAAAGTTTATCATGATTCAGCA 493  
QY 1632 ATACAGTCAATGCTTCAAGGCTCACTTCACTTTTAAATTCAGTCTTGTGTTCTAC 1691  
D 494 ATACAGTCAATCTTCAAGGCTCACTTCACTTTTAAATTCAGTCTTGTGTTCTAC 548  
QY 1692 CACATCTGCTGCTTCTTCTGATGAGTCAATCTTGAACCTTCAAGTCAATGAGGCT 1751  
D 549 CACATCTGCTGCTTCTTCTGATGAGTCAATCTTGAACCTTCAAGTCAATGAGGCT 608  
QY 1752 T-GAATCGACTTCTTCAAAATTCCTGTTAATTTATTTTGA---CCTCCATGAAT 1807  
D 609 TGGATATCAATCTTCTTCAAAATTCCTGTTAATTTATTTTGA---CCTCCATGAAT 668  
QY 1808 CATGAATGTTCTTATGATGACACC 1829  
D 669 CACGAATGTTCTTATGATGACACC 690

RESULT 7  
LOCUS HSM806813 9266 bp mRNA linear HTC 22-SEP-2004  
DEFINITION Homo sapiens mRNA; cDNA DKFZp686M21107 (from clone DKFZp686M21107).  
ACCESSION BX640738  
VERSION BX640738.1 GI:34364830  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
1. (bases 1 to 9266)  
Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R.,  
Mewes, H.W., Weill, B., Amlid, C., Osanger, A., Fobo, G., Han, M. and  
Wiemann, S.

CONSRPT  
TITLE The German cDNA Consortium  
JOURNAL Direct Submission  
Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuherberg, GERMANY

COMMENT  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email [s.wiemann@dkfz-heidelberg.de](mailto:s.wiemann@dkfz-heidelberg.de);  
sequenced by DKFZ (German Cancer Research Center).  
Heidelberg/Germany) within the cDNA sequencing consortium of the  
German Genome Project. This clone (DKFZp686M21107) is available at  
the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in  
Berlin, Germany. Please contact RZPD for ordering:  
<http://www.rzpd.de/cgi-bin/products/ci.cgi?cloneID=DKFZp686M21107>  
Further information about the clone and the sequencing project is  
available at <http://mips.gsf.de/projects/cdna/>.

## FEATURES

source

1. 9266  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="RZPD:DKFZp686M21107"  
/db\_xref="taxon:9606"  
/clone="DKFZp686M21107"  
/tissue\_type="fetal kidney"  
/clone\_lib="686 (synonym: hlcc3). Vector pSport1\_Sfi; host  
DH10B; sites SfiIA + SfiIB"  
/dev\_stage="fetal"  
/note="hypothetical protein, C-terminus elongated"  
1. 9266  
/gene="DKFZp686M21107"

gene

CDS

406..3903  
 /gene="DKEZP686M21107"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="CAB5851.1"  
 /db\_xref="GI:34364831"  
 /translation="MEODRTNHEGNRLSPILPSPICOTEPATLKLONGSPLEBA  
 HPEVNDGDKMHSFQYQPCMGKSGNSRVSPPDSRGSKLQNGQIKRTYSE  
 LSGLIQIKLKODKANGERRNRSVSOERNPSSQWVDDLSKESVSVAQENAV  
 KDTFSSTNCSGPNELQILNEQEGSANYDKNTVLNKNVLPNPNATVASAM  
 EHTHSELLEKTLQYDPQVSIQVQKTSIHNANSTQELSCETIHPHSTQINS  
 AQTNSLPRKPAVSEACDADADNASKLAMLNCSERLOKPELOQOQSVFICPS  
 PAENNIQGTTLASGEFSGSSSNLOAPGSSSRVLYKQEMNAYKQSVFLFED  
 SATTPPSPOLISPPPLPOVPLPSEKSTNGVLEHNNHVPKOSVTLREK  
 IEKREAPRPSQSPRSTHVSQSPMLSERQNNVNRNDQTAGTMTVPLCSETRFM  
 SEHLKHPPLFGSGELQDNCQOLMRNEDEILKRDKEQTRDVLPTQYVLEKWE  
 LKAPRFOQASHLKRNEASLPSILQYQPNLSNMTSQYQNSMPGALPROAYTKT  
 TOLEHKSQYOVEMNQSQGVDOHLQPKPSHOVHSTKIDLPKMHVOSTGTRH  
 FOORADQTEKLSPVLKQOLNQOASETEPPNSHLLQHPKHAQOTOPSLSHLPQ  
 NOOQOKLOIKKEELIOTPPHPSNNDQREGSPROQTVRECHENQYSSSSE  
 THVOMKEEVQINIRNSPSQTMKSSACKIOVSCSNTHLVSENEQTHPELFG  
 NKTQNLHMQYFPNNVLPKODLHRCFOQEOBQSQASVLQYKRNQDMSGQAO  
 AQORYLHNHANEVPEVDQGSHTQTPQKOTQKHALRMLQKQOQOQPOTES  
 CHSQMHPPIVEBQCKPHACMHTAPENKTKVTKYKQENPASCNDYQKSIETMO  
 HLKOPHAKSLFDKALTKSQOKQYKVMGSPVTLTQTTAAELDSHTPALBOOTTS  
 EKPTRTASVIANPRTSPSKLIDPIKMLDTPVTVQYDFSCRCVGCQCKTEPH  
 GYVELANLSSDMGFSFFP"

ORIGIN

Query Match 18.0%; Score 360.6; DB 4; Length 9266;  
 Best Local Similarity 67.6%; Pred. No. 2.6e-47;  
 Matches 783; Conservative 0; Mismatches 284; Indels 92; Gaps 16;

Db 868 CTCCTAATTTTATGCTATGCTATAATACATTAAGTTTGTGTTGTTTGTACCT 927  
 4333 CTCCTAAGTTTCCTTTCTCATTATTAATGGAATGTTGATTTATATCATTTT 4392

Qy 928 ATTGCTGTGGCTGGGTCAGCAACATTTTCTGTAAGGGCTAGATGACGGCATC 987  
 4393 GATTCCTCTTTCT---CTAATAATTCATCTTTTGTATTAAGAAATATGATACGACGATAC 4449

Qy 988 CTTGAGATACAGTGGGTTGGTTCCATACACACACATATATCAATATGCAAGAAGTG 1047  
 4450 CTTAGAGATATTTGGGTTGGCTCCATACACATATAA-----TG 4491

Db 1048 GATATACAAATTAAGTACACACAAAGCTTTTGGCTCCAGTGCATATAAAGTTT 1107  
 4492 AATATATACATTAAGCAAGTTGTAAGGACTTTTGGTTCTCACTGATGATAAAGTTAT 4551

Qy 1108 GCTTATATCACTGTAGTCTGTTAAGTGCATATAGTTTATGTTCTTAAAAACACATA 1167  
 4552 TTATATACTATACCTGTAACATCTAATGTCGCAATACATATGTCCTAAAAA-----ATA 4606

Db 1168 CTTAATTTTAAATGCTTATTAATCTTAAAAATGCTAACAATATTTGACATTCAGTGA 1227  
 4607 TATACCTTAAAAATTAATTTATGTTAAAAAAATGCCAACATTAATCTGGGCTTTAGTGA 4666

Qy 1228 GTTGTATCTTTTGTGTGTGAAGCTCTTTTCTTATGATGACTGATCGGAGGCTC--- 1283  
 4667 GTGCTATCTTTTGTGTGTGAAGGCTCGTCTGATATTGATCGGTGAGCATATCA 4726

Db 1284 -----AGTGTGGAAGCTTAAGGCTGTGTGCAAGTTTCTTAAAAAC 1327  
 4727 TGTGTGATGTTGCTGAAGGTTGCTGAGATGCTGTGTGTGCAATTTCTTAAAAATAG 4786

Qy 1328 AGTGAAG---ATTGCAATATCACTGATCTCTCTTTCATGAAGAATTTCTCTAGTGT 1384  
 4787 ACAACGTGAAGGCTGATCAATTAATTTTTCATTCACAAAAGATTTCTCTAGCAT 4846

Qy 1385 GTGATGCTTTTATAGCATTTTATGACAGTGAACCTTTTGAANAATGGA--TCATC 1443  
 4847 GCAATGCTGTTGATGATTTAACCAACAGCAAGAAATTTCTTGAANAATGGAATGAGTGC 4906

Qy 1444 CTCTCAAAACCCTGCTGCTTTTAACAACCTAAGTAAATATATATTCGAATTCATGTT 1503  
 Db 4907 CTCTCAACCTGCTGCTGCTTTTATCACTAAGTTTGTATATATTCGAATTCATGTT 4966

Qy 1504 GTCATTTCAACAAATTTTCAAGTGTCTTCAACAGAGTAGAATTCATTCATTCCTGAG 1563  
 Db 4967 GTCAATTTCAACAGTTT--ACAGATCTTCAATGGAATATATTCATCTC----- 5014

Qy 1564 ATGGAATCTTTGCTCATCCATTAAGAGAAATTCCTCATCTGTTCAAGTTTATCATAGA 1623  
 Db 5015 AAACATCTCTTTGTTCAATCAACAAAGAACAA-----CTTCTATCAAGTTTTCATAGCA 5069

Qy 1624 TTGAGCAATACAGTATGCTTTTCAAGGCTCACTTCATCTTAAATTCAGTCTCTGCT 1683  
 Db 5070 TTGAGTAACTCAAGCCCATCTTCAGS-----CTTACTTAAATTTGTTGTTCTGCT 5124

Qy 1684 GTTCTTACCAATCTGTGTTCTTCTTCCATTAAGTCTTGAACCTTCCAAATCAATCC 1743  
 Db 5125 ACATCTCCCATCTGCAAGTACCT--CTCAAGGAAGCTTGAATCTCTCAAGTATTC 5183

Qy 1744 ATGAGGTTGGAATCGACTTCTTCAAAATCTGTTAATATTAATTTTGA---CCTCC 1800  
 Db 5184 ATGAGGTTGGAATCAACTCT---AAACTCTGTTAATGTTATATTAAGACCCCTCC 5240

Qy 1801 CATGAATCAGAAATGTTCTTAATGACACCTGGAATGTTGAATCTTCCAAAAGTTTC 1860  
 Db 5241 CATGAATTAAGAAATGTTCTTAATTAATCTCT--AAATGTTGATCTTTCAGAAAGCTTTC 5299

Qy 1861 AATTTACTTATGTCAGATCCATCCATCCAGAGATCCATTTCAATGCCAGTTATAGCT 1920  
 Db 5300 AATGTAATTTGCCCCGATTCATC-----AGAAACTATCTTGGCAGCTGTAAGT 5349

Qy 1921 TATGAAATGATTTTCTTCAATTAATTAAGCTTGAAGTTGAATTAATCTTGAATTCAT 1980  
 Db 5350 AA--CAATATATCTCTTAATTAATTAAGACTTGAAGTGAAGTCAATCTTAATTCATAG 5407

Qy 1981 TCTGCAAAATAGATGTTGT 1999  
 Db 5408 GCTGCAAGATCAATGTTGT 5426

RESULT 8  
 LOCUS CB957195/c 778 bp mRNA linear EST 29-APR-2003  
 DEFINITION AGENCOURT\_1366972 NIH\_MGC\_184 Homo sapiens cDNA clone  
 CB957195  
 ACCESSION CB957195  
 VERSION CB957195.1 GI:30213312  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 778)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE Unpublished (1999)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: sgabbs-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: NDCM153 row: p column: 10  
 High quality sequence stop: 637.  
 Location/Qualifiers  
 1..778  
 /organism="Homo sapiens"

## FEATURES

source

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3053361"
/lab_host="DH10B (71 phage-resistant)"
/clone_1lb="NIH_MGC_184"
/note="Organ: Pooled-glandular; Vector: pDNR-LIB; Site: 1:
SfiI (ggcgcctcgcc); Site 2: SfiI (ggcgcctcgcc);
Library is oligo-dT primed and directionally cloned. cDNA
was prepared from a glandular pool of tissues from thyroid,
parathyroid, adrenal, cortex and pineal gland. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCATTATGACC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCCGAGCGCCGACATC-dT(30)BN-3' (where B = A,
C, G or N = A, C, G, or T). Average insert size 1.38
kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC library."

```

## ORIGIN

```

Query Match      18.0%; Score 360.4; DB 6; Length 778;
Best Local Similarity 76.5%; Pred. No. 4.8e-47;
Matches 525; Conservative 0; Mismatches 141; Indels 20; Gaps 6;

1320 AAAAAAGAGTGAAGATTCAGATTCAGTACTCTCTTTCATGAGATTTCTCT 1379
      |||||
676 ACAACAACATGAGATTTGTCGATCGATTCCTCCACAAAAATTTTCTGT 617
      |||||
1380 AGTGTGATGCTTTTGTATGATGATTTTATGACAGTAACTTTTGAATTTGA 1438
      |||||
616 AGCATGAGTCTATTGTGTAGCATTTTACCACAGAGAACTTTCAAAATTTAG 557
      |||||
1439 CAATCTCTAAACCCGCTCTGCTTAAACAAGTAAATATA--TATCTGATC 1496
      |||||
556 CGATACTGTAACCTGCGACCTGTTATCACTTATTAATTAATGACTTAACT 497
      |||||
1497 CATTTGTGATTTCAACAATTTTCAAGTGTCTTCCAGAGATAGATTCATCTCAT 1556
      |||||
496 CTTTGTGATTTCAACAATTTTCAAGATCTTCCAGAGATTCATCTCAT 437
      |||||
1557 TCTGAGATGAACTTTGCTCATCAATGAAGAAATTCCTCATCTGTTCAAGTT 1616
      |||||
436 AAACCACTT---TCTTGTTCATCCTTAAGAACAACTCCTCACTTCAATTT 381
      |||||
1617 CATGAGATTCAGAAATACAGTCTATGCTTCAAGGCTTCATCTTAAATTCAG 1676
      |||||
380 CATGAGATTCAGAAATTCAGTCTATCTTCAAG---CTCCACTTCTAATTTGA 326
      |||||
1677 TCTTGTGATTTCAACAATTTTCAAGTGTCTTCTCAATGAACTTTCCTCAAA 1736
      |||||
325 TCTTACTATTTCACCAATTCAGATTTATTTCTTCACTGAAGTCTTGAACCT 266
      |||||
1737 GTCAATCATGAGGTTGAAATGCACTTCTTCAAAATTCCTGTAAATTTATTT 1794
      |||||
265 GTTATTCATGAGGTTGTAATCACTTCTCAAAATTCCTGTATGTAATTTGAC 206
      |||||
1795 -ACCTCCATGATCAATGATTTCTTAATGACACCTGATGGATTCCTTCCAAA 1853
      |||||
205 CACCTCCATGATCAATGATTTCTTAATGATGTAATGATTAATGATTCCTTCC 146
      |||||
1854 GGTTCATTTACTTGTAGTCAGATTCATTCATCCAGAGATTCATCTTCAATG 1913
      |||||
145 AGTTTCAATGATCTTGTCCAAATGAT---CAGAGATCACTATCTATGAGAG 91
      |||||
1914 ATAGCCTTATGAAATGATTTCTTCAATTAATGAGCTTGAAGTGAATTAATCT 1973
      |||||
90 ATAGCCTTATGAAATGATTTCTTCAATTAATGAGCTTGAAGTGAATTAATCT 31
      |||||
1974 TCCATTTTCTGCAAAATGATGTTGT 1999
      |||||
30 TCCATGGGCTGAGGAAAAAGATGCTGT 5
      |||||

```

```

RESULT 9
BG113167/c      994 bp      mRNA      linear      EST 30-JAN-2001
LOCUS          602283837F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4371176 5',
DEFINITION    mRNA sequence.
ACCESSION     BG113167
VERSION       BG113167.1 GI:12606673
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE     1 (bases 1 to 994)
              NIH-MGC http://mhc.nci.nih.gov/.
AUTHORS      CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMIL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clome distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/ILML at:
              http://image.llnl.gov
              Plate: L1AM10029 row: j column: 09
              High quality sequence stop: 637.
COMMENT       Location/Qualifiers
              1..994
                location="Homo sapiens"
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4371176"
                /tissue_type="osteosarcoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_1lb="NIH_MGC_86"
                /note="Organ: Bone; Vector: pCMV-Sport; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.53 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC library."

```

## FEATURES

Source

1..994

```

/location/Qualifiers
1..994
  location="Homo sapiens"
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:4371176"
  /tissue_type="osteosarcoma, cell line"
  /lab_host="DH10B (phage-resistant)"
  /clone_1lb="NIH_MGC_86"
  /note="Organ: Bone; Vector: pCMV-Sport; Site 1: NotI;
  Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
  Average insert size 1.53 kb. Library enriched for
  full-length clones and constructed by Life Technologies.
  Note: this is a NIH_MGC library."

```

## ORIGIN

```

Query Match      17.6%; Score 351.6; DB 2; Length 994;
Best Local Similarity 77.5%; Pred. No. 1.1e-45;
Matches 535; Conservative 0; Mismatches 129; Indels 26; Gaps 8;

1257 TTTCTTATGATGATGATGAGGAGGCTGCTGATGAACTTAAAGTGGCTGTG 1316
      |||||
675 TTGATGATTTCTTGATGATGATGAGGAGGCTGCTGAAAGTGGGCTGTGCA 616
      |||||
1317 CTTTAA---ACAAAGTGAAGATTCAGATTCAGTT---GACTCTTCTTTATG 1369
      |||||
615 CTTAAATGAAGACATCGTGAAGCTTTTGCATGAGGTGACTGTCTGTTATGA 556
      |||||
1370 ATTTCTCTAGTGTGATGCTTTTGTATGATTTTATGACAGATGAAGTCTTTGA 1429
      |||||
555 ATTTCTCTGATGATGATGAGTGTGATGATTTTGGCCACAGTAGAGCTTTCA 496
      |||||
1430 AAATTTGA--TCAATCTCTCAAAACCTGCTGCTTTTAAACAAGTAAATTA 1488
      |||||
495 AAATTTGAGTCAATCTCTCAAAACCTGCTGCTTTTAAACAAGTAAATTA 436
      |||||
1489 TCTGAATCATGTTGTCATTTTCAAAATTTTCAAGTCTTCAAGAGTGAATTC 1548
      |||||
435 TCTAAATCTTTTGTTCATTTTCAAAATTTTCAAGTCTTCAAGAGTGAATTC 376
      |||||
1549 ATCTCATTTCTGATGATGATCTTTGCTCATCATCAATGAAGAAATTCATCT 1608
      |||||
375 ATCTCAAGAAACCACTT---TCTTTGCGATTCATTAAGAAACCACTCTCAT 320
      |||||
1609 AGTTTATCATGATGATTCAGCATATGATGATGATGATGATGATGATGATG 1668
      |||||

```



[illegible]

QY	1115	CTACACGTGTAGCTGTTAAAGTGCAGTAAGCTTATGTCTTAAAAA---CACATACCTT	1171
Db	62	CTAATCTGTAGCTATT- TGGGTGCATATGCATTAATGTCTTAAAAACAATGTAAATACCTT	120
QY	1172	AATTTTAAAAATGCTTATTTACTTAAAAATGCTAAACAATCAT- TTGAGCATTTCAAGTAGT	1228
Db	121	AATTAAAAAATACTATATTTGTCTTAAAAAGATGTCTAACATCATCGGTAGCCCTTCAAGTAGT	180
QY	1230	TGTAAATCTTTTGGCTGTGGAAGCTCTTTTCTT-----ATTGATGACTGATCGGG	1279
Db	181	AGTAATCTTTTATGCTGGCAGAGGGGCTTGCCTTATGTGTGATGTAGCTGTGACTGATCAGA	240
QY	1280	GGTACAGGTCGAAGCTTAAGGGTGGCTGTGCAGATTTCTTAAAAACAAGGAAGATGTC	1339
Db	241	ATGTGTTTGGTGAAGCTTGGCTGGCTGCACGCAATTTGTGAACACAATGAAGTTTGC	300
QY	1340	AATATCAGTGAATCTTCTCTTTCATGGAAGATTTCTCTGTAGTGTGATGCTTTTGAT	1399
Db	301	CTCCTGTATGATTTCTTCTTTCACGAAAGATTTCTCATGTAGCATGCTGTGTTGAT	360
QY	1400	AGCATTTTATGACAGTGAACCTTCTTTGAAAATTGCA- TCAATCTCTCAAACTCTGCT	1458
Db	361	AGGGTTTATCCACAGATTAATCTTCTTCAAAAATGGAGTTGATCTCTCGAATCTGCT	420
QY	1459	CTGCTTTAACAACCTAAGTTAATPAAATATTTCTGAATCATATGTTGTGATTTCAAC- AAT	1511
Db	421	GCTGCTTATCAACAAGTTTATTAAGATTTTCTAAACCTTTGTGTATTTCAACAAT	480
QY	1518	TTTTCACAGTGTCTTCCACGAGAGTAGATTCATCTCATTTCTGTAGATGTAATCTTGTCT	1577
Db	481	TTTTCACAACATCTTGCCAGAGATAGATTCAGAGTCAAGAACACACTT---TCTTTATTT	536
QY	1578	CATCATTAAGAAGAAATTCCTCATCTGTTCAGTTTATCATGTGATTTGCAGCAATACAG	1637
Db	537	TAACCATTAAGAAGCAACGCTCATTTCAATTGAAGTTTATCATGTGACGTGCAGCAAGCTAG	596
QY	1638	TCATGTCTTTCAGGCGTCACTTCACTTTAAATTTTCAGTTCTCTGCTGTTCTTACACATC	1697
Db	597	GCAATCTTGAAGGCTCCACTT-----ATTCTAAATTTCTTTGCTGTGCTTACCATATTC	648
QY	1698	TGTGATTCCTTCTCTCATATTGAAGTCTTGAACCTCTCCAGATCATCATGAAGGTTGAAT	1757
Db	649	TGCAGTTACTTCTTCTCTGTAAGTCTTGAACCTCCCTCAAAAGTTATTCATGAGGTTAAAT	708
QY	1758	CGACTTCTTCCAAATTCCTGTTAATATTTAATTTTGAACCTCC	1800
Db	709	CAACTTCTTNCAAACTCTGTGTATGTGTCTATTATTTTAACTTCC	751
RESULT 11			
BX955959/c			
LOCUS	BX955959	719 bp	mRNA
DEFINITION	DKFZP781A1775_r1 781 (synonym: hlc4) Homo sapiens cDNA clone		linear EST 01-MAR-2004
ACCESSION	DKFZP781A1775.5		mRNA sequence.
VERSION	BX955959		
KEYWORDS	BX955959.1	GI:43439531	
ORGANISM	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Ctenidata; Vertebrata; Euteleostomi;		
ORGANISM	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
ORGANISM	Homidae; Homo.		
REFERENCE	1 (bases 1 to 719)		
AUTHORS	Pouskka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,		
AUTHORS	Mewes,H.W., Well,B., Amdt,C., Osanger,A., Fobo,G., Han,M. and		
AUTHORS	Wiemann,S.		
AUTHORS	ESR (Pouskka,A., Albert,R., Moosmayer,P., Schupp,I.,		
AUTHORS	Wellenreuther,R., et al.)		
AUTHORS	Unpublished (2003)		
COMMENT	Contact: MIPS		
COMMENT	MIPS		
COMMENT	Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany		











```

OY 1718 AAGCTTGAACCTCTCCAG----TCATCATGAGGTTGGAATGACTTCTTCCAAAT 1773
DB 70 AAATCTTGAGACCCCTCAAGGGGTTGCTCATGAGGTTGAATGACTTCTCCAAACT 11
OY 1774 CCTGTAAAT 1782
DB 10 CCTGTAAAT 2

RESULT 17
CK774525 831 bp mRNA linear EST 20-FEB-2004
LOCUS 963293 MARC 2BOV Bos taurus cDNA 3', mRNA sequence.
DEFINITION CK774525
ACCESSION CK774525
VERSION CK774525.1 GI:42728668
KEYWORDS EST
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Bovidae; Bovinae; Bos.
1 (bases 1 to 831)
Smith,T.P.L., Grose, W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Caass,E., Wray,J.E., White,J., Cho,J., Fahrnkung,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chicko-McKown,C.G., Perlea,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)

JOURNAL 11282978
PUBMED Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
plate: 97 row: C column: 14
Seq primer: GTAATACGACTCCTACTATAGGG.
FEATURES
source Location/Qualifiers
1..831
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DHL0B"
/clone_id="MARC 2BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN
Query Match 16.4%; Score 328.2; DB 7; Length 831;
Best Local Similarity 71.2%; Pred. No. 5.6e-42;
Matches 606; Conservative 0; Mismatches 193; Indels 52; Gaps 11;

OY 1179 AAATGCTTATTAATAAAAAATGCTACATCATTTGAGCATTGAGTGTGTAATCTT 1238
DB 1 AAATACTTATTAATCAAAACCTGCTAACCATCATTTGGCGATTGAGTGCATATATCTC 60
OY 1239 TTGCTGTGGAAGTCTTTTCTTATTTGATGACTGAT-----CGGGGGTCAGGT 1287
DB 61 TTGCTGTGGAAGTCTTTTCTTATTTGATGACTGAT-----CGGGGGTCAGGT 120
OY 1288 GCTGAAGTTAGGGTGGCTGGTGGCAGTTCTTAAACAA-----CAGTGAAGAT 1336
DB 121 GCTGAAGTGGGGTGGTGGAGGCAATTTCTTAAATTAAGCAATAGGCAATGAGTT 180

```

```

OY 1337 TGCAATATGAGTTGACCTCTTCTTCAATGAAGAATTTCTCTGAG-----TGTGATGC 1391
DB 181 TGCGGCATCAACCTTCTTTCTTCCATGAACAATATTTGTGAGCATGATGCAATGC 240
OY 1392 TTTTGTATGACATTTTATGACAGATGAACCTCTTGAATAATGG-ATCAATCCCTCAA 1450
DB 241 TGTTTAGTGACATTTATTCACAGAGAGCTTCTTCAAAATTTGAATCAATCTCTCA 300
OY 1451 ACCCTGCTGTGTTTAAACAACCTAAGTAATATATATTTGATTCATTTGTTCATT 1510
DB 301 ACCCTGCACTGCTTTATCACTAAGTTATGTAATATTCATTAATCTTGGTCATT 360
OY 1511 CAACAATTTTACAGAGTCTTCCACAGAGATGATTCATCTCATTTCTGATGAGAT 1570
DB 361 CAACCATCTTACATGATCTTTCACAGAGATGATTCGCTCA-----AGAGCCAC 413
OY 1571 CTTTGTGATCATGAAGAAGAAATTCCTCATCTGCAAGTTTATCATGATGACGAC 1630
DB 414 GTTCTCTGCTGACCGGTGAAGAGCTCTCAGCATTAAGTTTATCATGATGACGAC 473
OY 1631 AATACAGTATGCTTTCAGGCTCCTCACTTCACTTTAATTCAGTTCTTGTGTTCTA 1690
DB 474 AACTCACTCCACTGAGGCTTCACCTT-----TAATCTAGCTGTGCTCTTCA 528
OY 1691 CCACATCTGTGCTCTTCTTCCATTTGAAGTCTTGAACCTCTCCAGATCATGAGAG 1750
DB 529 TCACATCTGATGATCTTCTTCCACAGAGTCTTGGCCCTTCAAGTCAACCATGAGGA 588
OY 1751 TTGAATGACCTTCTTCCAAATTTCCGTAAATTAATTTATTTGACC-----TCCCATGAT 1807
DB 589 CTGGAATCAACTTCTTCCACACTCCGCTGTAAGTTGATTTGACCTTTCTATGAT 648
OY 1808 CATGATGTTCTTAAATGACCTGAGATGATGATCTTTCACAAAGTTTCAATTTAC 1867
DB 649 CACAAATGTTCTTAATGACATCTGAATGATGATCTTCC-AGAGTTTCAATTTAG 707
OY 1868 TTAGTCCAGATCCATCCATCCAGAGATCCATCTTCAATGCCATGATTAAGTGA 1927
DB 708 TTGGCCAGATTCAT-----CAGTGAATTAATTAATCTTGGCAGCTATAACCTAAATGGA 762
OY 1928 TGATTTCTTCAATTAATGAGCTTGAAGTTGAA-TTACTCTGATGATCATTTCTGCA 1986
DB 763 --ACATATTTCTTAAATTAAGACTTGAAGGTTGAATTTAGCCTTATTCATGACTGCA 820
OY 1987 AAATGATGTT 1997
DB 821 GAAATGATGTT 831

RESULT 18
CR859436/c 5336 bp mRNA linear HTC 12-NOV-2004
LOCUS CR859436
DEFINITION Pongo pygmaeus mRNA; cDNA DKFZp469P156 (from clone DKFZp469P156).
ACCESSION CR859436
VERSION CR859436.1 GI:55729758
KEYWORDS HTC.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Pongo.
1 (bases 1 to 5336)
Wandut,R., Heubner,D., Mewes,H.W., Weill,B., Amid,C., Oeanger,A.,
Fodo,G., Han,M. and Wiemann,S.
The German cDNA Consortium
CONSTRM Direct Submission
JOURNAL Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuberberg, GERMANY
REFERENCE Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
sequenced by Agowa (Berlin/Germany) within the cDNA sequencing
consortium of the German genome project.
This clone (DKFZp469P156) is available at the RZPD Deutsches

```



/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:464319"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1ib="Nih\_MGC\_20"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 16.3%; Score 326.6; DB 2; Length 929;  
Best Local Similarity 78.3%; Pred. No. 9,86-42;  
Matches 474; Conservative 0; Mismatches 109; Indels 22; Gaps 6;

1386 TGATGCTTTTGTATGATTTATGACAGTAGAAGTCTTTGAAATTTGATCATC-- 1443  
|||||  
899 TGAATCTGTTGTATGATTTATGACAGTAGAAGTCTTTGAAATTTGATCATC-- 840  
1444 -CTCTCAAAACCTGCTCTGCTTTACAACTTAATTAATTAATTTGATCATC-- 1501  
839 TCTCTAATCAATGCACTGCTTATGATTAATTAATTAATTTGATCATC-- 780  
1502 TTGTCAATTTCAACATTTTCAAGTGTCTTCAACAGAGATGATTCATCTCCTG 1561  
779 TTGGCATTGCAACAAATTTCAAGCTTCTTCAAGAGATGATTCATCAAGTACCA 720  
1562 AGATGAATCTTTGTCATCATCAAGAAATTTCCATCTGTTCAAGTTTATCATCA 1621  
719 TG-----TTTGTGCTATCATTAAGAAATTTCTCCATCTGTTCAAGTTGATCA 665  
1622 GATTGAGCAATCAAGTATGTCATGCTTCAAGGCTCACTTCACTTTAATTTCAAGTCT 1681  
664 GATTGAGCAATCAAGTATGTCATGCTTCAAGGCTCACTTCACTTTAATTTCAAGTCT 610  
1682 CTGTTTCTACCAATCTGCTGCTTCTTCCATTTGAAGTCTTGAACCTTCCAGTATC 1741  
609 CTAGTTCGCGCACATCTGCAAGTCTTCTTCCATGAACTTGAACCTTCCAGTATC 550  
1742 CCATGAGGTTGGAATGAGCTTCTTCAAAATTTCCGTTAATTTATATTTG--ACCT 1798  
549 CCATTAAGGTATGATCACTTCTTCAAACTTCTGTTAATTTGAAATTTGGCTTCT 490  
1799 CCATGATCATGATGTTCTTAAATGCACTGGAATGGAATCTTTTCCAAAGGTTT 1858  
489 CCATGATCATGATGTTCTTAAATGCACTGGAATGGAATCTTTTCCAAAGGTTT 430  
1859 TCAATTTAATTTAGTCCAGATTCATCCAGAGATCACTTTCAATGCAAGTATATGC 1918  
429 TTAATTTAATTTAGTCCAGATTCATCCAGAGATCACTTTCAATGCAAGTATATGC 375  
1919 CTATATGAATTTATTTCTTCAATATATAGGCTTGAAGTTGAATTTAATTTAATTT 1978  
374 TTTATGAATTTATTTCTTCAATATATAGGCTTGAAGTTGAATTTAATTTAATTT 315  
1979 TTTCT 1983  
314 GGGCT 310

RESULT 20  
LOCUS CE040574 C 755 bp DNA linear GSS 24-SEP-2003  
DEFINITION tigr-gss-dog-1700035006572 dog library Canis familiaris genomic,  
genomic survey sequence.  
ACCESSION CE040574  
VERSION CE040574.1 GI:35072379

KEYWORDS GSS.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Carnivora; Placentalia; Canidae;  
Canis.

REFERENCE 1 (bases 1 to 755)  
Kirkness E.F., Bafna V., Halpern A.L., Levy S., Remington K.,  
Ruch D.B., Delcher A.L., Pop M., Wang W., Fraser C.M. and  
Venter J.C.

TITLE The dog genome: survey sequencing and comparative analysis  
JOURNAL Science 301 (5641), 1898-1903 (2003)  
PUBMED 14512627

## COMMENT

Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

## FEATURES

source

Location/Qualifiers

1..755  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_1ib="dog Library"  
/note="Site 1: Becki; Libraries were prepared from peripheral blood"

## ORIGIN

Query Match 16.3%; Score 325.6; DB 9; Length 755;  
Best Local Similarity 72.7%; Pred. No. 1,56-41;  
Matches 528; Conservative 0; Mismatches 164; Indels 34; Gaps 7;

1277 GGGGTCAGGTGCTGAAGCTTAGAGTGGCTGTCGAGTTTCTTAACAACAGTGAAGAT 1336  
751 GGGTGGCGGTGTGGACAGGTGGGTGGGATTTCTTAATGTAAC--ATGTT 696  
1337 TGCATATCAAGTTGACTCTTCTTATGAAAGATTTCTCTAGTGTGATGCTTTT 1396  
695 TGGTGATGATTTAGGCTTCTTCAATGATCAATTTCTGTCGATGCGGTCGTT 636  
1397 GATAGCTTTTATGACAGTGAAGCTTTTGAATTTGATCAATCTCTCAAACTCTG 1456  
635 GAAAGCAATTTTACACAGTGGACGTTTCAATTTGAAATGATCAATCTCTCAAACTCTG 576  
1457 CTGCTTTTAAACAACCTTAATTAATTAATTTCTGAATCAATTTGTCATTTCAACA 1516  
575 TAGACGTTTATCAACCAACATATGTAATTTTAATTTCTTGGCATTTCAACAA 516  
1517 TTTTCAAGTGTCTTCAACAGAGTATGATTCATTTCTTCCGAATGGAATCTTGC 1576  
515 TCTTACAGACCTTCAACAGAGAAATTTTATCCCA--AGAAACCATGTTATTCG 461  
1577 TCAATCAATGAAGAAATTTCTCATCTGTTCAAGTTTATCATGATGATGAGATACA 1636  
460 TCATTTTAAGAACCACTTCACTTGAAGCTTATATAGATTT----- 412  
1637 GTCAATGCTTCAAGGCTTCACTTCAATTTCAAGTTCTTCTGCTGTTTCAACAT 1696  
411 GTCACTTTTCAAGGCTTCACTTCA--AATTCAGTTCTTCTGCTGTTCTACTGAT 357  
1697 CTGTGTTCTTCTTCAATGAAGCTTGAAGCTTCAAGTCAATGATCATGAGGTTGAA 1756  
356 CTGATGTTACTTCTTCACTG-AGTCTGAAGCTTCAAGTCAAGTCAAGTATGTTGAA 298  
1757 TCGACTTCTTCAAAATTTCTGTTAATTTATTTATTTGAC--TCCATGATCATGAA 1813  
297 TCACTTTTCTCAAACTCTGTTAATTTGATTTTACCACTTTTCAATGATGCAAG 238  
1814 TGTTCTTAATGCACTGGAATGGAATCTTTTCAAAAGTTTCAATTTAATTTAGTTC 1873

Db 237 TGTCTTACCGACACTGATAGTGGACTCTTCCAGAAAGTTTTCACCTTGTGT 178  
QY 1874 CAGATCCATCATGATGAGATCCATTTGAAAGCCAGTTATAGCCTTATGATGAT 1993  
Db 177 CAGATTCAT-----CAGAAAGATCATCTGTATGGAAGCTGTACCTTACCTAATGAT 123  
QY 1934 TCTTCATATTAAGAGCTGGAAGTTGAATTAATCTCTGATGCCATTTTGTGCAAAATAGA 1993  
Db 122 TCTTAATATATATATGATGATGATTAATTAATTAATCTCTGATGCCATGAGGCTGCAATGGA 63  
QY 1994 TGTGTGT 1999  
Db 62 TGTGTGT 57  
RESULT 21 815 bp DNA 1linear GSS 21-APR-1999  
AO237450/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
1 (bases 1 to 815)  
Adam, M.D., Rounsley, S.D., Zhao, S., Baas, S., Linher, K., Golden, K., Berry, K., Granger, D., Sub, E., Wible, C., de Jong, P. and Venter, J.C.  
Use of human BAC End Sequences for Sequence-Ready Map Building  
Unpublished (1998)  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
BACPAC Resources (http://bacpac.med.buffalo.edu). Clones may be purchased from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1..815  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7523340"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-61M13"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_lib="RPCI-11"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"

Db 724 TTATATACATAGTCACTATTAATATGATGATGATGCTGTAAATAACATGCA 665  
QY 1164 CATACCTTAATTTTAAATGCTTATATACAAAAATGCTAACATATATTTAGCATTC 1223  
Db 664 CATACGTTAATTTAAAGATACCTTTATGCTTATATGCTAACATATCATGCGCTTCA 605  
QY 1224 GTAGTGTAAATCTTTTGTGCGGAAAGCTC-----TTCTATTTGATGATGA 1274  
Db 604 GCAATATCATATCTTTTGTGCGGAAAGTTTGTCCCTCGATATGCTGCTGATGATCA 545  
QY 1275 TC-GGGGGTCAAGGTGCTGAAGCTTGAAGGTGCTGAGCAGTTCTTAAACAACTGAA 1333  
Db 544 TCAGGGGTGCTTGTGCTGAAGGTGCTGAGGTGCTGAGCAGTTCTTAAATAAC---AA 489  
QY 1334 GATTCGAATATCAGTTGACCTTCTTCTTATGAAAGTTTCTCTATGCTGATGCTT 1393  
Db 488 GTTTGGCCCATCAATGACCTTCTCTCTGAAAGCTTCTGTAGCCTGCGAAGCTG 429  
QY 1394 TTTGATAGCATTTTATGACAGTGAAGCTTCTTGAATATGATCAATGCTGCAAGC 1453  
Db 428 TTTTATATGACATTTTGGCCGAGTGAAGTTCTTCAAAA---GTCAATCTCTCAACT 373  
QY 1454 CTGCTGCTTCTTAAACACCTTAATATATATATTTCTGAATCATTTGTGATTTCAA 1513  
Db 372 CTGCTACTGCTTAA-----CCAATATATGTAATATTTTAAATCTT---CATATTTCA 320  
QY 1514 CAATTTTACAGTGTCTTACCAAGAGTGAATTCATTTCTGATGATGAAATCTT 1573  
Db 319 CAATATCTACAGCATCTCAACCAAGTGAATCACTCATCAGAT---AACTAGTCTT 264  
QY 1574 TGTCTATCATTAAGAAGAAATTCATCTGTCAGTTTATCATGATGATGACAGAT 1633  
Db 263 TGTCTATCATTAAGAAGAAACCAACTATCTTCAAAATTTTATCATGATGATGACAGAT 204  
QY 1634 ACAGTCATGCTTCAAGGCTCACTTCACTTTTATTCAGTTCTTGTGCTTTTACCA 1693  
Db 203 GATGTCATCTTCAAG-----ATTCAATTCATTAATACATTTTGTATTTTCCACCA 149  
QY 1694 CATCTGTGCTTCTTCTTCATTTGAAGCTTGAACCTCTGCAATCATCATGAGGTTG 1753  
Db 148 TATCTG-----CCTTCTCATTTGAATCTTGAATCCCTCAAGTCATCATGAGGTTG 94  
QY 1754 GATCAGCTTCTTCAAAATTCCTGTAATATTTATTTTATGAC---CTCCCATGATCAT 1810  
Db 93 GATTAACCTTCTTCAAGCTTTTGAATGTTGATTTTGAATCTTCTTGAATCAT 34  
QY 1811 GAATGTTCTTAATGCACTGGAATGGAAT 1842  
Db 33 GAATGTTCTTAATGCACTGGAATGGAAT 2  
RESULT 22  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
1 (bases 1 to 705)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgaabs-rt@mail.nih.gov  
Tissue Procurement: James Martin  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa



cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@iowa.edu  
The following repetitive elements were found in this cDNA  
sequence: 1-38, >AT-rich#low complexity 39-705,  
>TIGER1/DNA/MER2\_type (matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

SOURCE

Location/Qualifiers

1..705

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-H-FH1-Dfu-m-19-0-UI"

/tissue\_type="Cell line"

/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="NCI\_CGAP\_FH1"

/note="Organ: Chondrosarcoma; Vector: pT73-Pac  
(pharmacia) with a modified polylinker; Site: 1: EcoR I;  
Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library  
obtained from a cell line derived from grade I  
chondrosarcoma tissue. The library was constructed and  
normalized according to Bonaldo, Lennon and Soares, Genome  
Research, 6:791-806, 1996. First strand cDNA synthesis was  
primed with an oligo-dT primer containing a Not I adaptor.  
Double stranded cDNA was ligated to an EcoR I adaptor,  
digested with Not I, and cloned directionally into  
pT73-Pac vector. The oligonucleotide used to prime the  
synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
AGAACCGGC. The cell line was provided by Dr. James Martin  
from the University of Iowa.  
TIG-TTSUB=Human Chondrosarcoma Cell line CS8 - Grade 1  
Chondrosarcoma  
TAG\_LIB=UI-H-FH1  
TAG\_SEQ=AGAAATCCGCGC"

## ORIGIN

Query Match 16.3%; Score 325.4; DB 5; Length 705;  
Best Local Similarity 75.4%; Pred. No. 1.6e-41;  
Matches 522; Conservative 0; Mismatches 132; Indels 38; Gaps 8;

974 TAGTACGAGCATCCTTGAGATCTGTGGTGTGTTCCATACACACACATATACAA 1033  
|||  
34 TAAATACGGCATACCTCAAAAGATATGTGATTTGGTCCAGACCATGTAAT----- 87  
1034 ATATGCAAGAAGTGATATCAATAAAGTACACACAGCTTTTGGCTTCCAGTG 1093  
88 -----AAGGCAATATCATATAATAAGTACCAACTTTTGT- TTCCAGTG 137  
1094 CATATAAAGTTTGTCTTATATCACTGTAGTCTGTTAAGTGTGCAATAGTATATGTC 1153  
138 AGTATAAAGTTATGTTTACACTATACAGTCTATTAAGGTTCAATTAACATTATGTC 1197  
1154 TAAAAAA-ACACATACCTTATTTTAAATGCTTATTAATAAAAAATGTAACATCAT 1212  
198 TAAAAAATGTAATACCTTAAATTAATAATCTTATTAACAAAAATACATGATCAT 257  
1213 TTGAGCATTCAGTATGTTATCTTTTGTGCTGGAAGTCT-----TTTCTT 1262  
258 CTAAGCCTTCAGTGGTGTGATCTTTTGTGGGGAAGGCTTGCTGCTGATGTGATG 317  
1263 ATTGATGATGATCGGGGCTCAGGTCTGAAGCTTAAGGCTGCTGTGCACTTTCTTAA 1322  
318 GCTGCTGATGATCAGAGAGTGTGATGCTGAAGGCTAGGAGGCTGTGCAATTTCTTAA 377  
1323 -----ACACAGTGAAGATTCGAATATGATGATCTCTCTTCATGAAAGATTTCTCT 1377  
378 ATAAACAAACATGAAGTTTGTGCTGATCAATTTGACTTTCCTTTCATGAAGATTTCTCT 437

QY 1378 CTAGTGTGATGATGCTTTTGTATGACATTTTA-TGCACATGTAAGAACTTTTGAATTTG 1436  
|||  
DB 438 GTAGTGTGATGATGCTTTTGTATGACATTTTAACCATATGTAAGAACTTTTGAATTTG 497  
QY 1437 A-TCAATCTTCAAAACCTTCTGCTTTTAACAAACCTTAATATATATCTGAAT 1495  
DB 498 AGTCAGGCTCTCAAAACCATCCACTGTTTATCACTAATTTATGTAATATTCAAAT 557  
QY 1496 CCATTTGTGATTTCAACAAATTTTCAAGTGTCTTCCAGAGATAGATTCATTCAT 1555  
DB 558 CTTTGTGCTCAATTTCAACAGATGATCACTGTCTTCCAGAGATAGATTCATTCATCA 617  
QY 1556 TTCTGATGATGAATCTTTGCTCATCTCAATGAAGAAATTTCTATCTGTTCAAGTTTA 1615  
DB 618 GAACACACTT-----TCTTTGTCATCAATTAAGAAACAATCCTCATTCATTCAGTTTGA 673  
QY 1616 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1647  
DB 674 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 705

RESULT 23  
AQ748474 844 bp DNA linear GSS 19-JUL-1999  
LOCUS  
DEFINITION  
HS 5540 A2 E04 T7A RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=1116 Col=8 Row=1, genomic survey sequence.  
ACCESSION  
AQ748474 GI:5535632  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euteria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo  
1 (bases 1 to 844)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,W.D. and  
Hood,L.

REFERENCE  
AUTHORS  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,W.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 1116 row: I column: 8

Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 844.  
Location/Qualifiers

## FEATURES

SOURCE

1..844

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="Plate=1116 Col=8 Row=1"

/sex="male"

/clone\_lib="RPCI-11 Human Male BAC Library"

/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"

## ORIGIN



Query Match	16.3%;	Score 325.2;	DB 9;	Length 844;
Best Local Similarity	75.9%;	Pred. No. 1.7e-41;		
Matches 486; Conservative	0;	Mismatches 133;	Indels 21;	Gaps 6;

OY		1369	AATTTCTCTGAGTGTGGATGCTTTTTGATTGCAATTTTANGCAGTAGAAGACTTC	1428
Dd		15	GAA TTC TCT G TAG CTA GGA A AT G CT GT AGA T B GA C AT TT TA CC CA G TA G AA C TT C TT TC	74
OY		1429	AAAAATGGA-TCAATCTCTCAAAACCCGTCTGCTTTAACAACCTTAATATAATA	1487
Dd		75	AAAATTGAGA-CAAATCCCTTTCAAAACCTTGCTGCTCTTATATCAATAGTTACTGTATA	134
OY		1488	TTCGAATTCATATGTTGTCATTTTCAACAATTTTCAAGTGTCTTACCGAGATGATTC	1547
Dd		135	TTCTAAATCTTTTGCGCATTTTCATTAAGTGTTCACACATCTTCACAGTAGATGATTC	194
OY		1548	CATCTCATTTCTCTGATGAGATCTTTTCTCATCATTAAGAAGAAATCCATCTGTTTC	1607
Dd		195	TGTCTCAAGAGGCCACTT---TTT TTTTCTCATCATTAAGATATCTGTTCATATGTTA	251
OY		1608	AAGTTTATCATATGAGATTGCAGCAATACATGATCATCTTTCAGGCTCATCTTCACTTTAA	1667
Dd		252	AAGTTTATCAATAGATATTTCAGCTATTAAGTCAACATCTTCAG-----CCTTCACCTTCCA	306
OY		1668	TYCAGATCTCTGCTGTTTCTTATCACACATCTGTGTGTTCTTCTCATTTGAAG---TCT	1723
Dd		307	TTCTTATTTCTCTTGCTCTTTTCCACCAATCTGCAAGTATCTTCTTCACTGAAGTCTTCT	366
OY		1724	TGAACCTCTCCAGATCATCATGAGGTTGGAATGCACTTCTTCCAAATTCCTGTATA	1783
Dd		367	TGAACCCCTTGAAGCATCTTGAAGGGGTGGAATCACTTCTTCAAAATGCTCTTAATC	426
OY		1784	TTTATATTTTGA---CCATCCATGAATCATGAATGTTCTTATAGGCACCTGGAATGGTGA	1840
Dd		427	TTGATATTTTGAACCTCTCCCAACATATATGAATGTTCTTATAGGCATCTTGAAGGGTGA	486
OY		1841	ATCTTTTCCAAAAGTTTTCATATTACTTACTTACGTCCAGATCATCATCCAGAGATCCACT	1900
Dd		487	GTCTTTTCCAGAAGTTTTCATTTACTTTGTGTGAGCTGAGCTGAG---CAGAGATCAACC	541
OY		1901	TTCAATGCCATTTATAGCTTATAGAAATGATATTTCTTCAATATATAGGCTTGAAGTTGA	1960
Dd		542	ATCATATGCACTATGCTTTTACAAAAGTATTTCTTGAATATATGAAGCTTGA AAAATGGA	601
OY		1961	AATTAATCTCTTGATCATCTTTTCTGCAAAATATAGATGTTGTG 2000	
Dd		602	AATTAACCCCTTGACCCATGCGGCTGCAAGTTGAATTTGTG 641	
RESULT 24				
CC478317/c				
LOCUS				
DEFINITION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

## ORIGIN

Query Match	16.2%;	Score 325;	DB 9;	Length 849;
Best Local Similarity	70.1%;	Pred. No. 1.8e-41;		
Matches 581;	Conservative	0;	Mismatches 205;	Indels 43;
			Gaps	9;

Qy	1043	AAGGAGATATACACATTAAGAGAGTCCACAAAGCTTTGGCTTCCGAGTCATTA	1102
Db	811	AAGCAATTCATTAATTAAGAGAGTCCACATGAATTTTGGTTCCAGTCATTA	752
Qy	1103	GTTTGTCTTATCTACACTGTAGTCTGTTAAGTGTCAATAGTGTATGTC-TA	1161
Db	751	GTATGTATTACACTATCTAGTGTATTAAGTCCACAACGATTAATGTC	692
Qy	1162	CACATACCTTAATTTTAAATNGCTTTATTAATAAAATGCTAACATTTAG	1222
Db	691	TACATTCCTTAATTAATAAATTTCAATGTTAAAAATGCTAACCTGATGAG	632
Qy	1222	CAGTAGTTGTA---TCCTTTGCTGGTGAAGTCTTTCTT-----ATTG	1268
Db	631	TAAACAATTGACATCTTTTTTTTTTGTTGGAGGTCCTGCTGATGTAGT	572
Qy	1269	GACTGATCGGGGGCTCAGTGTCTGAAGCTTAG---GTGGTGGCAGTTCTT	1322
Db	571	GACTGATCAGTGTGTAGTTCCTCAAAATCGGGTAGTGTGTGGCAATTTCT	512
Qy	1325	AACAG---TGAAGATGCAATATCAGTGTGACTCTCTTCATGAAGAATTC	1381
Db	511	AAGATATGAAGTTTCTCCATGATGACTCTTCTCTTCAGAGCAATTTCTT	452
Qy	1382	TGTGTGATGCTTTTGATAGCATTTTATGCAAGTAGAATCTTTGAAATG	1441
Db	451	CATGCAATGCTGTTTGACAGCATTTAAACCAATAGAACTTTCAAAATTA	392
Qy	1442	TCTCTCAAAACCTGCTCTGCTTAAACAACCTAAGTTAATATATTTCTGA	1501
Db	391	TCTCTCAAAATCGATGTGCTTTCACACTTAGTGAATTAATTTCTAAAT	332
Qy	1502	TTGTGATTTCAACAATTTTCACAATGTCTTCAACAGAGTAGATTCATCT	1561
Db	331	TTGTGATTTCAACAACCTTCAACAATCTTCAACAGAGTAGATTTAGCTCA	272

QY 1562 AGATGAATCTTGGTCATCCATAGAGAAATTCATCTGTCAGATTATCATGA 1621  
DB 271 ACTT-----TTTGTGTCACCCCATAGAGCAAGTTTCAATTAATCTGTATCAGCA 216  
QY 1622 GATTGACAGCAATACAGTATGCTTCAGGCTTCACCTTCATTCATTATTCAGTTCTTTG 1681  
DB 215 TATTCACAGCAATTCAGACATATCTCCAGGCTCCAC-----TTTCAGTTATCTTG 167  
QY 1682 CTGTTTACACACATCTGTGTCTCTCTCCATTCATGAACTTTGAACCTCTCCAAAGTCAT 1741  
DB 166 CTATTCGCTGCATCTGGGGGTACAGCTGCCACGCAAGCCCTTGAGCCCTCAAAAGTCAT 107  
QY 1742 CCATAGAGGTTGGATTCACCTTCTTCCAAATTCCTGTTAATTAATTAATTTGACCT--- 1798  
DB 106 CCATAGAGGCTGAGAGAA---CTTCCAACTCTGTTATGATGATTTGGCTCTT 50  
QY 1799 -CCCATGAATCAATGTTCTTAATGACACCTGGAATGATGATCTT 1846  
DB 49 CCCCATGAATCTGAATGTTCTTAATGACGTCTAGATGATGATGATCTT 1

RESULT 25  
B0067736 1039 bp mRNA linear EST 02-APR-2002  
LOCUS B0067736.1 GI:19896782  
DEFINITION AGENCOURT\_6643057 NIH\_MGC\_121 Homo sapiens cDNA clone IMAGE:5768547  
5', mRNA sequence.  
ACCESSION B0067736  
VERSION B0067736.1 GI:19896782  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgsaps-r@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LHAM2828 row: j column: 04  
High quality sequence stop: 655.  
Location/Qualifiers

FEATURES  
source  
1. 1039  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5768547"  
/lab\_host="DH10B"  
/clone\_id="NIH\_MGC\_121"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI;  
Site 2: EcoRV (destroyed); RNA source anonymous pool of 3  
fetal brains, female age 20 weeks, female age 24 weeks,  
and male age 26 weeks. Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.7 kb, insert size range  
0.7-3.5 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 017. Note:  
this is a NIH\_MGC Library."

## ORIGIN

Query Match 16.2%; Score 324.2; DB 3; Length 1039;  
Best Local Similarity 70.4%; Pred. No. 2.3e-41;  
Matches 589; Conservative 0; Mismatches 208; Indels 40; Gaps 10;

QY 1059 AAGTAGCTACACAAAGCTTTTGGCTTCCAGTGCATATATAAGTTTGGTTATACCTAC 1118  
DB 829 AAGGCAATCACACAAATCTTTTGGTTGGCTGGCAATAGAACTGCCGTATATCAATAT 770  
QY 1119 ACTGATGCTGTAAAGTGTGCAATAGCTTATGCTAAAAAACAATACCTTAATTTTA 1178  
DB 769 ATAGTAGCTTAATAGGTATACGA-CGGCATATGCTAAAAATGATATACCTTAATTTAG 711  
QY 1179 AAATGCTTTATATCAAAAAAATGCTATACAAATCATTTGAGCATTCAGTAGTTGTA----- 1234  
DB 710 AATATTTATTTATCAAAATTTATTTAATATATATGAGCCCTGAATGATGCTTTTGTG 651  
QY 1235 -----TCTTTTGTGTGAGAGGCTTTTCTAT-----TGATGACTGATCGGG 1279  
DB 650 TTGATGTTGTTGTTGTTGATGATGAGTGTGCACATATAGTGGCTGCTGATGATCA-G 592  
QY 1280 GGTGAGGCTGTAACCTTAGGTTGGCTGTGGAGTTTCTTAA-----ACACAGGAG 1334  
DB 591 GGTGATGCTGTAAGGTAGGGGTGGCTGTGGCAATTTCTTAAAGTACACAAAAATGATG 532  
QY 1335 ATTGCAATTCAGTTGACTCTTCTCTTCATGAAAGATTTCTCTAGTGTGATGCTTT 1394  
DB 531 TTTATCAATCAATTAATGACTCTTCTTTAGAAAAGTTTGTGTGGACACAAATATCTGT 472  
QY 1395 TTGATAGCATTTTATGACAGTAGAATCTTTGAAAATGATGATCAATCTCTCAAAACC 1454  
DB 471 TTGATAGCATTTTACCATGAGAAAGCT--TTCAAAATGGAATCAATCTTTGCAAAATCC 414  
QY 1455 TGCTGTGCTTTAAACAACCTTAATTAATTAATTTCTGAATCATTTGTCATTTCAAC 1514  
DB 413 TGCTGTGCTTTATCAACTTCAAT--ATGACATATTTCTTAATCTTTTGTCTATTTTAA 355  
QY 1515 AATTTTACAGATGCTTTCACAGAGTAGATTCATTCATTCATTTCTGAGATGAAATCTTT 1574  
DB 354 AATTTTACAGAGCTCTTTCACAGAGTAGATTTCTGCTCAAGAAACACTT-----TCTTT 299  
QY 1575 GCTCATCATTAAGAAAGAAATTCCTCATCTGTTCAAGTTTATCATAGATGACGAATA 1634  
DB 298 GCTCATCATTAAGAAAGAAATTCCTCATCTGTTCAAGTTTATGATGACGAATAAT 239  
QY 1635 CAGTATGCTTTAGAGCTTCACTTCTTTAATTCAGTCTTGTGTTCTTCAAC 1694  
DB 238 CTTTGCATCTTCAAG-----CTACACTGCAATTTAGTTCTTTGCTATTTTCAAC 184  
QY 1695 ATCTGAGTCTTCTTCTTCCATTCATGAGTCTTGAACCTCTCAAGTCAATCAAGAGGTTG 1754  
DB 183 ATCTGCTGTTAATCTCTTCTTCTAAGATTTGAAGCCCTCAATCTCAAGAGGTTGT 124  
QY 1755 AATGCACTTCTTCAAAATTCCTGTTAATTTATAT--TTGACTTCCATGATCATGA 1812  
DB 123 GATCAACTTCAATCAAAATCTCTGAATAATATCATATCTATCTCCATGATCATGA 64  
QY 1813 ATGTTCTTAATGCACTGGAATGATGATCTTTTCCAAAGGTTTCAATTTACTT 1869  
DB 63 ATGTTCTTAATGATGCTGATGATGATGATGATTTTCCAGAGGTTTCAATTTACTT 7

RESULT 26  
A0750123/c 853 bp DNA linear GSS 19-JUL-1999  
LOCUS A0750123 HS 5575 A2 C03 SP6 RPCI-11 Human Male BAC library Homo sapiens  
DEFINITION genomic clone Plate=1151 Col=6 Row=E, genomic survey sequence.  
ACCESSION A0750123  
VERSION A0750123.1 GI:5537281  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo  
1 (bases 1 to 853)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

TITLE  
Hood, L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
PUBMED  
10449764  
COMMENT  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Research Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 1151 row: E column: 6  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 853.  
Location/Qualifiers

## FEATURES

1. .853  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="plate=1151 Col=6 Row=E"  
/sex="male"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/note="Vector: pBACe3.6; site 1: EcoRI; site 2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACe3.6 vector at EcoRI sites"

## ORIGIN

Query Match 16.1%; Score 321.4; DB 9; Length 853;  
Best Local Similarity 73.8%; Pred. No. 6.6e-41;  
Matches 558; Conservative 0; Mismatches 136; Indels 55; Gaps 8;

1043 AAGTGATATCAATTAAGTAGAGTACACAAAGCTTTGGCTCCAGCATATATAA 1102  
691 AAGTGAATATCGAATAAAGCAAGTAACACATTTT-----TTCTAGGCATATATAA 636  
1103 GTTTTGCTTAACTACAGTGAAGTCTGTTAAGTGCATAAGTGTATGTTAAAAAAC 1162  
635 GTTAGGTTTACACTCTTCTAGTCTATTAAAGTGT----- 599  
1163 ACATACCTTAATTTTAAATGCTTTATTAATAAAGCTAACATATTGAGCATTC 1222  
598 ACATACCTTCAATTTTAAATGCTTTTAAATGCTAACATATTGAGCATTC 542  
1223 AGTGAAGTGAATCTTTTGTGCTGAGGAGCTTTTCTTA-----TTGATGATC 1272  
541 AGTGAAGTGAATCTTTTGTGCTGAGGAGCTTTTCTTA-----TTGATGATC 482  
1273 GATCGGGGGTCAAGGCTGAGCTTGAAGTGTGCTGAGGAGCTTTCTTA-----ACAAC 1327  
481 GATCGGGGGTGTAGCTGAAGATGTGCTGAGGAGCTTCTTAATAAAGCAAC 422  
1328 AATGAAGATGAATACAGTGAAGCTTCTTCAATGAAGATTTCTCTAGTGTG 1387  
421 AATGAAGATGTGAGTGAAGTGAAGCTTCTTGAATAAAGATTTCTCTAGTGTG 362  
1388 ATGCTTTTATGATGATTTTATGACAGTGAAGCTTTTGAATAATGGA-TCATCTC 1446  
361 ATGCTTTTATGATGATTTTATGACAGTGAAGCTTTTGAATAATGGA-TCATCTC 302  
1447 TCAAAACCTGCTCTGCTTTAACAACCTTAAGTAAATTAATTAATTCATTCATGTTGTC 1506  
301 TCATCTCTGCTCTGCTCTTCACTCACTAAGTTATGATTAATTCATTCATTCATGTTGTC 242

QY 1507 ATTTCACAATTTTACAGTGTCTTCACGAGATGATTCATTCATTCCTGAGATG 1566  
DB 241 ATTTCACAATTTTACAGTGTCTTCACGAGATTCATTCATTCATTCCTGAGATG 183  
QY 1567 GAATCTTTGCTCATTCATTAAGAAGAAATTCATCTGTTCAAGTTTATCATGAGATG 1626  
DB 182 -TTTGTGCTAACCCGTAAGAAGCAATCTTCATTAAGTTGATGATGAGATG 126  
QY 1627 CAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1686  
DB 125 CAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 71  
QY 1687 TCACACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1746  
DB 70 TCACACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11  
QY 1747 AGGTTTGA 1755  
DB 10 AGGTTTGA 2

## RESULT 27

BG028427 931 bp mRNA linear EST 24-JAN-2001  
LOCUS BG028427/c  
DEFINITION 602294332F1 NIH\_MGC\_86 Homo sapiens cDNA clone IMAGE:4389342 5',  
mRNA sequence.  
ACCESSION BG028427  
VERSION BG028427.1 GI:12417521  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS NIH-MGC http://img.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-r@mail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLINL at:  
http://image.llnl.gov  
Plate: L1AM10076 row: 0 column: 07  
High quality sequence stop: 681.  
Location/Qualifiers

## FEATURES

1. .931  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4389342"  
/tissue\_type="osteosarcoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 86"  
/note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally, oligo-dT primed.  
Average insert size 1.533 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 15.9%; Score 318; DB 2; Length 931;  
Best Local Similarity 74.6%; Pred. No. 2.2e-40;  
Matches 491; Conservative 0; Mismatches 135; Indels 32; Gaps 6;  
QY 1360 TTCATGAAGATTTCTCTTCTAGTGTGATGCTTTTGTATGATGATGATGATGATG 1419  
DB 697 TGCCTTCAGAGAGATTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 638



QY 1701 GGTTCTTCTCCATTGAAGTCTTGAACCTCT 1732  
 DB 662 AGTACTTCGTCCACTGAAGTCTTGAACCCCT 693  
 RESULT 29  
 CFI26937/c 703 bp mRNA linear EST 05-AUG-2003  
 LOCUS UI-HF-ETO-avx-m-05-0-UI.r1 NIH MGC\_214 Homo sapiens cDNA clone  
 DEFINITION  
 CFI26937  
 CFI26937.1 GI:33204674  
 EST.  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 703)  
 Bonaudo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 JOURNAL  
 PUBMED 8889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Mary Hendrix  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/humanfl.html  
 The following repetitive elements were found in this cDNA  
 sequence: 1-701, >TIGERI#DNA/MER2\_type  
 Seq primer: PYX-5.  
 FEATURES  
 source  
 1..703  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30563524"  
 /tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH MGC 214"  
 /note="Organ: Lung; Vector: PYX-Asc; Site: 1: EcoR I  
 Site 2: Not I; The library was constructed according  
 Bonaudo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction. ligated  
 with EcoR I adaptor, digested with Not I and then cloned  
 directionally into PYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GATTAAGGCCA. Tissue was provided by Mary Hendrix."

ORIGIN  
 Query Match 15.8%; Score 315.8; DB 6; Length 703;  
 Best Local Similarity 73.4%; Pred. No. 5.2e-40;  
 Matches 509; Conservative 0; Mismatches 152; Indels 32; Gaps 7;  
 QY 1076 TCTTTGGCTCCAGTCATATAAAGTTTCTTACTACACTGTAAGTCTGTAAGT 1135  
 DB 703 TATTTTATTTTCCATACCTAATAAAGTTATTTAATCTATATATAGTCTATTAAGT 644  
 QY 1136 GTGCATATAGTGTATGCTTAAAAAACAATACCTTAATTTTAAATGCTTATTAATCTAA 1195  
 DB 643 TTGCATATACATATATCTGACAAATGTACATTAATTAATAAATATTTATTTCTAA 584

QY 1196 AAAATGCTAACATTCATTTTGAGCATTCAGAGTTGTAATCTTTTTCGTGGAAGTC 1255  
 DB 583 AAGATGCTAACATTCATTAAGTCTTCAAGT-----CTTGTCTGGAACAGTC 529  
 QY 1256 TTTTCTTATGAGTACTGAT-----CGGGGCTCAGTGTCTGAGCTTAAGTGGTGG 1304  
 DB 528 TTGGCTACAGTTGATGGCTGCGTAATGATCGGTGTGTGTGTGTCTGAAGTTAGTGTGG 469  
 QY 1305 CTGTGGCAGTTTCTTAA-----ACAACTGTAAGATTGAATATCACTGACTCTTCTT 1359  
 DB 468 CTCTAGAGATTTCTTTTAAATGAGCAACAAGAGGCTGCAACATCACTGACTCTTCTT 409  
 QY 1360 TTCAATGAAGATTTCTCTAGTGTGTGATGCTTTTGTATGACATTTTATGACAGATGA 1419  
 DB 408 TTC-CGAAAGATTATCTGTAGAGATGTTGCTGTGTGATGAGCTTTTATCCACAGTGA 350  
 QY 1420 ACTTCTTTGAAATATGGA-TCAATCTCTCAAAACCTGCTCTGCTTAAACACTAAGTT 1478  
 DB 349 ACATCTCTCAAAATTTGAAGCAATCTCTCAAACTGCTGCTGTTTATCACTAAGTT 290  
 QY 1479 AATATAATTTCTGAATTCATGTTGTGTCATTTCAACAATTTTCAAGTGTCTTACAGG 1538  
 DB 289 TATGTAAATATTTAAGTCTTGTGTGTCATTTCAACAATGTCACAGATCTTACAGG 230  
 QY 1539 AGTAGATTCATCTCATTTCTGAGATGAATCTTGTCTATCATTAAGAAATTCCT 1598  
 DB 229 AGTAGATTCATCTCAAGAACCACTT---TTGTGCTCTTATATTAAGAACCACTCT 174  
 QY 1599 CATCTGTCAGATTTTATCATGATGATGAGCAATTCAGTCAATGCTTCAAGGCTCACTT 1658  
 DB 173 CATTTGTCACGTTTATATCATGATGATGAGCAATTCAGTCAATGATGAGG-----TTC 119  
 QY 1659 CACTTTTAATTCAGTGTCTTGTGCTGTTTCTACACATCTGTGTTCTTCTCAATGA 1718  
 DB 118 CACTTCTAAGTCAAGTATCTTCTGATTTTCCACACATCTGCACTGATCTTCTGCACTGA 59  
 QY 1719 AGTCTGAAACCTTCCAGATCATCATGAGGCT 1751  
 DB 58 AGTCTTGAACCCCTCAAGATCATCATCAGGCT 26

RESULT 30  
 CG734253/c 822 bp DNA linear GSS 20-OCT-2003  
 LOCUS RP11-57117, T7 RPCI-11 Human Male BAC Library Homo sapiens genomic  
 DEFINITION  
 clone RP11-57117, genomic survey sequence.  
 CG734253  
 ACCESSION CG734253.1 GI:37776745  
 VERSION  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 822)  
 Zhao, S. and Yonescu, R.  
 Direct SubMISSION of BAC End Sequence of CCAP clones  
 Unpublished (2003)  
 COMMENT Contact: Yonescu, R  
 NCI/NIH  
 9000 Rockville Pike, Bethesda, MD 20892, USA  
 Tel: 301-402-2008  
 Fax: 301-402-1204  
 Email: yonescu@pop.nci.nih.gov  
 Plate: 57 row: 1 column: 17  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 source  
 1..822  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

```

/clone="RP11-57117"
/sex="male"
/clone_lib="RP11-11 Human Male BAC Library"
/notes="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"

```

```

ORIGIN
Query Match      15.7%; Score 314.8; DB 10; Length 822;
Beet Local Similarity 69.3%; Pred. No. 7.3e-40;
Matches 592; Conservative 0; Mismatches 212; Indels 50; Gaps 10;

QY 979 CAGGATACCTTGGAGATCTGTGGTGGTTCATACACCAATATATCAATATG 1038
DB 822 CAGACATTCCTCAAAATATATGACAGGGGTTCCAGTCCAGACACCAATATATG 763
QY 1039 CAAAGAGTGTATATCAATATTAAGTACACAGCTTTGGCTCCAGTGCATAT 1098
DB 762 ACAACACATGA-----GTCTATGATATATATCTCTCAGAGAGTCT 717
QY 1099 AAAAGTTTGTCTTACTACATCTGTAGTCTGTTAAGTGCATATGTTATGCTTAA 1158
DB 716 AAAAGTTTGTCTTACTATATGTTATGTTAAGTGCATATGTTATGCTTAAAC 657
QY 1159 AAAACATACCTTAATTTTAAATGCTTTATTAATAAAATGCTTAAATCATTTGAGC 1218
DB 656 AAAGCATATCTTAA-----AAAAAGTTTATGATATATATATATATATATCTGA-C 602
QY 1219 ATTCAGTGTATATCTTTTGTCTGTGTGAGAGGCTTTCTTATTTAGTACGTAT--- 1275
DB 601 CTTCATATATTTTAACTTTTGTGTGTGAGAGGCTTTCTTAAAGTGTAGGCTGCT 542
QY 1276 -----CGGGGTCAGGTGTGAAGCTTAGGGTGGCTGTGCAAGTTTCTTAAACAC 1327
DB 541 GACTGATCAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 482
QY 1328 AGTG-----AAGATTGCAATATCAGTTGATCTTCTTTCATGAAAGTTTCTCTTATG 1383
DB 481 ACAGCATATTAATTTGCTGTATGATGATGATGATCTTCTTCAAAAAGATT---TTTCTGCA 424
QY 1384 TGTGATCTTTTGTATGATGATTTATGACAGATGAACTTCTTGAATTTGATGATCA 1443
DB 423 TGTGAATGCTTTGATTAACATTTTAACTTC---TGAATGCTTTCAAAATTTGAGTTAAT 367
QY 1444 CTCTCAAAACCTGCTGCTTAAACAATTAATTAATTAATTAATTAATTAATTAATTA 1503
DB 366 CTTCCCGATATCTGCTGCTTATATCACTTAAGTTTGTATATTTCTGTATCTTTGTG 307
QY 1504 GTCTATTTCAACATTTTCAAGTGTCTTACACAGAGATGATTCATCTCATTTCTGAG 1563
DB 306 GTCAATTTTCAAAATGCTATGATCAATCAACAGAGATGATTCATCTCATTTCTGAG 250
QY 1564 ATGAATCTTTGCTATCATATGAAGAAATTTCTCATCTGTGCAAGTTTATGATGA 1623
DB 249 CCACTTTCTGCTGATTCATGAGAGCACTCTCATCTGTGCAAGTTTATGATGAAT 190
QY 1624 TTGAGCAATACAGTCAATGCTTCAAGGCTCATCTTCACTTTAATTTCAAGTTCTTGT 1683
DB 189 TTGTCGCAATTCAGTCAAGTCTTCAAGG-----CTTCACTTTAATTTGATTTCTTGTG 135
QY 1684 GTTCTTCAACATCTGTGTTCTTCTTCAATGAAAGTCTTGAACCTTCTCAAGTCAATCC 1743
DB 134 ATTTTCAACATCTTCAAGTATTTCTTCCGCTGAAGTCTTGAACCTTCAAGTCAAGC 75
QY 1744 ATGAGGGTGAATGACTTCTTCAATTTCTGTTAATTTAATTTAATTTGA---CTCC 1800
DB 74 ATGAGGGTGAATTAATCTTCAACAACTCTATTAATTTGATATTTTAAAGCTTCTTC 15
QY 1801 CATGAATCATGAAT 1814
DB 14 CATGAATCATGAAT 1

```

```

RESULT 31
BU740723      668 bp  mRNA  linear  EST 10-OCT-2002
LOCUS         UI-E-EU0-ai1-h-12-0-UI.s1 UI-E-EU0 Homo sapiens cDNA clone
DEFINITION   UI-E-EU0-ai1-h-12-0-UI 3', mRNA sequence.
ACCESSION    BU740723
VERSION      BU740723.1 GI:23681014
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 668)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL      PUBMED
COMMENT      Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 16-652, >ITIGER1#DNA/MER2_type (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

```

## FEATURES

source

```

1..668
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EU0-ai1-h-12-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev stage="fetal and adult"
/lab host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EU0"
/notes="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoRI; Site 2: Not I;
UI-E-EU0 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoRI adaptor, digested
with Not I, and cloned directionally into pRTT3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes,
AGATTCAGAG; lens, CGATTAGGCA; eye anterior segment,
ATGACCGCAT; optic nerve, CCATTAGAGG; retina, CCGCG; Retina
Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
library was created for the program, Gene Discovery in the
Visual System, supported by National Eye Institute (NEI).
TAG_Lib=UI-E-EU0
TAG_SEQ=CCGCG"

```

## ORIGIN

Query Match

15.6%; Score 312.6; DB 5; Length 668;





	Best Local Similarity 72.3%;	Pred.No. 2,2e-39;	Mismatches 566;	Conservative 0;	Mismatches 178;	Indels 40;	Gaps 11;
QY	1113 TACTACACTGTAGTCTGTGTAACTGATGTCATATGTTATGTCTTAAAAA---CAATACC	1169					
Db	774 TACACTACTGAGACTCTTTTAAATGTGCTACCTGGCATATGTATTAAAAACAATGTACATATC	715					
QY	1170 TTAAATTTAAAG-CTTATTTACTTAAAAAAGCTAAACAATCTTGAACATCTTCACTGAG	1228					
Db	714 TTAATTTTAAAAAGACTTTTACAAGTAAACTTCTAGAGATCATCTGAGCCTTCCAGCAAG	655					
QY	1229 TTGTATCTTTTGTCTGTGGAAGTCTTTCTTATGTATGATGAT-----CG	1277					
Db	654 TCATATATCTTTTGGCAGATGGAAGGGCTTCTCTCATGTATGTGTGTGTAAGTATCAG	595					
QY	1278 GGGGTCAAGTGTGTAAGCTTAGGTGGCTGTGCAGTTTCTTAAAAACAAGGAAGAT	1337					
Db	594 GGTGTGGTGTCTGAAGGTAGGGGTGTGTGCAATTTCTTAAATTA-AGCAACGAA	537					
QY	1338 GCATATACATGTTGACTCTTCTTTCATGAAGAATTTCTCTATGTGTGATGATCTTTTG	1397					
Db	536 GTTGGCCAGATGTAGCTCTCCCTTTCAGAAAGATTTCTGTGAGAGGCAATCTGTG	477					
QY	1398 ATAGCATTTATATGACAGATGAACTCTTTGAAATTTGATCATCTCTCAAAACCTGC	1457					
Db	476 ATAGCA-ATTACCCACAGTAGAATCTT-TTAAAAATGGAGTCAAGTCTCTTAAAAACCTGC	419					
QY	1458 TCTGCTTTAACAACTTAAGTTAATATATATTCGATTCATTTGTGTCAATTTCAACAAT	1517					
Db	418 TGCTGCTTTATCACTAAATTTATGAACATCTTAAATCTTTGTCAATCTTTCAGCAAT	359					
QY	1518 TTTACAGATGCTCTCACAGAGATGATTCATCTCATTTCTGAGATGAAATCTTGTCT	1577					
Db	358 GTTCCAGCTCTTCA-----GATTAAGATTCACATGAA-----AAAACTTTATTTGTCT	307					
QY	1578 CATCCATTAAGAAGAAATCTCTCATCTGTCTCAAGTTTATCATGAGATTTGACAGAAATCAG	1637					
Db	306 TAGCCATTAAGAAGCAATCTCTCATCTCAATTTGATTTATCATATGAGATTTGACAGAAATCAG	247					
QY	1638 TCATGCTTTACAGGCTCATCTCATTTAATTCAGATTCCTGTCTGTCTTCAACATC	1697					
Db	246 TCACATTTTCAG-----ATTCACTCTTAATTCAGATTCCTGTCTGTCTTATATACATC	192					
QY	1698 TGTGATCTCTTCCTTCATGTAAGTCTTGAACCTCTCAAGTATTCATAGGGTTGAAT	1757					
Db	191 TGCAG-----TTTCAACAACATGAAGCTTAAACCCCTCAAAATCATTTCTGAGGGCTGGAAT	136					
QY	1758 GCACTTCTTCCAAATTCCTGTTAATTTATTTATTTTGAATCTTCCCATGAAATCATAGTT	1817					
Db	135 CAACGTTTGCACAACTCTCTTAAAGCTGAATATTTTGACCTCC-----TCCAGGAATGTT	80					
QY	1818 CTTAATGGAAGCTGGAATGATGATCTTTTCCAAAAGTTTTCATTTACTTATGTCAGA	1877					
Db	79 CTTCAAGGATCTTAAATGTAATCTTTTCCAGAAAGTTTTCATTTACTTTCCAGA	20					
QY	1878 TCCATC 1883						
Db	19 TTCAATC 14						
RESULT 34							
LOCUS	BM715720	724 bp	mRNA	linear	EST 28-FEB-2007		
DEFINITION	UI-E-EJ0-abj-k-13-0-UI-r2 UI-E-EJ0 Homo sapiens cDNA clone						
ACCESSION	BM715720						
VERSION	BM715720.1						
KEYWORDS	GI:19028978						
SOURCE	EST.						
ORGANISM	Homo sapiens (human)						
	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;						
	Hominidae; Homo.						





```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686B0107"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="686 (synonym: hicc3)"
/notes="vector: pTriblex2; site_1: sf1a; site_2: sf1b;
cDNA-collection"

```

## ORIGIN

```

Query Match      15.6%; Score 311.4; DB 5; Length 673;
Best Local Similarity 73.4%; Pred. No. 2.6e-39;
Matches 504; Conservative 0; Mismatches 151; Indels 32; Gaps 7;

QY 1064 GAGTCACACAAGTCTTTGGCTTCCAGTGCATATPAAAAGTTTGTCTTATCTACACTGT 1123
DB 1 GAGTCACACAATTTATTTTATTTCCAGTACATPACATATGTTATCTATATATAT 60
QY 1124 AGTCGTATAGTGTGCATATAGTGTATGCTPAAAAAACACATACCTTATTTTAAATG 1183
DB 61 AGTCATATTAAGTTTCAATAGCATATATATCTACATGACATCTTATTTTAAAAATA 120
QY 1184 CTTTATATCTAAAAATGCTTACAAATCATTTGAGCATTCAGTGTATCTTTTGTGC 1243
DB 121 TTTTATTTGTAAGATGCTTACGATATCTAAGTCTTCAGTGAAT-----CTTGTGTC 175
QY 1244 TGGTGAAAGTCTTTTCTTATGATGATGAT-----CGGGGCTCAGTGTCTGA 1292
DB 176 TGGAGACAGTCTTGGCTCAGTGTGATGCTGATGATGATGATGATGATGATGATGAT 235
QY 1293 AGCTTAGGTGCTGCTGCGAGTGTCTTAA-----ACACAGTGAAGATTTGCAATACG 1347
DB 236 AGGTAGTGTGCTGCTGAGATTTCTTAAATGACACAGAGAGGCTGCAACATCA 295
QY 1348 TTGACTCTTCTTCATGAAAGATTTCTCTAGTGTGATGCTTTTGTATGATGATTT 1407
DB 296 CTGACTCTTCTCTTC-CGAGGATTTATCTGATGATGATGATGATGATGATGATGATGAT 354
QY 1408 ATGCAAGTGAAGCTTCTTTGAAAATTTGA-TCATCTCTCAACCCCTGCTGCTTTA 1466
DB 355 ACCCAAGTGAAGCATTTCTCAAAATTTGAAGTCAATCTCTCAACCTGCTGCTGCTTT 414
QY 1467 ACAACTAGTATATATATATCTGATGATGATGATGATGATGATGATGATGATGATGAT 1526
DB 415 ATCACTAGTATATATATATATCTGATGATGATGATGATGATGATGATGATGATGATGAT 474
QY 1527 GTCTTACCAAGAGTATGATTCATCTCATTTCTGATGATGAAATCTTTGCTCATCTAA 1586
DB 475 ATCTTACCAAGAGTATGATTCATCTCATTTCTGATGATGAAATCTTTGCTCATCTAA 530
QY 1587 GAAGAAATTCCTCATCTGTTCAAGTTTATCATGATGATGATGATGATGATGATGATGATGAT 1646
DB 531 GAAGCAACTCTCATCTGTTCAAGTTTATCATGATGATGATGATGATGATGATGATGATGATGAT 590
QY 1647 CAGGCTCACTTCACTTTTAAATTCAGTCTCTGCTGCTTTCTTACCAATCTGATGCTTC 1706
DB 591 CAGG-----TTTCACTTCTAAGTCTAGTATCTTGTCTATTTTCCACCAATCTGAGTTAC 645
QY 1707 TTCTTCATGAGTCTTGAACCTCTC 1733
DB 646 TTCTTCATGAGTCTTGAACCTCTC 672

```

## RESULT 36

```

CE455208      671 bp  DNA      linear  GSS 27-SEP-2003
LOCUS      CE455208
DEFINITION  tigr-gss-dog-17000319501415 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION   CE455208
VERSION     CE455208.1 GI:36749305
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris

```

```

REFERENCE
AUTHORS      1 (bases 1 to 671)
              Kirkeness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
              Rueden, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
              Venter, J.C.
TITLE        The dog genome: survey sequencing and comparative analysis
JOURNAL      Science 301 (5641), 1898-1903 (2003)
PUBMED      14512627
COMMENT      Contact: Kirkeness EF
              The Institute for Genomic Research
              Department of Eukaryotic Genomes, TIGR, 9712 Medical Center Drive,
              Rockville, MD 20850, USA
              Tel: 301-838-0200
              Fax: 301-838-0208
              Email: ekirkenes@tigr.org
              Class: Shotgun.
FEATURES
source      Location/Qualifiers
              1..671
              /organism="Canis familiaris"
              /mol_type="genomic DNA"
              /strain="Standard Poodle"
              /db_xref="taxon:9615"
              /clone_id="Dog Library"
              /note="Site 1: BstXI; Libraries were prepared from
              peripheral blood"

```

## ORIGIN

```

Query Match      15.6%; Score 311.2; DB 10; Length 671;
Best Local Similarity 72.6%; Pred. No. 2.8e-39;
Matches 510; Conservative 0; Mismatches 158; Indels 34; Gaps 7;

QY 1287 TGGTGAAGCTTAAAGGCTGCTGCGAGTCTTAAACACAGTGAAGATGCAATATCA 1346
DB 1 TGGACAGGTGGGTGGGTGGGCAATTTCTTAATGTAAAC-----ATGTTGGTGATTTG 56
QY 1347 GTTGACTCTTCTTCATGAAAGATTTCTCTAGTGTGATGCTTTTGTATGATGATTT 1406
DB 57 ATTTAGTCTTCTTTCATGATCATATTTCTCTGAGCATCGGAGCTGTTGAAAGCATTT 116
QY 1407 TATGCAAGTGAAGCTTCTTTGAAAATTTGATCATATCTCTCAAAACCCTGCTGCTTTA 1466
DB 117 TACACACAGTGGGACCTCTTCAGAAATGGAATCAATCTCTCAAAATCTGTAGACCTTT 176
QY 1467 ACAACTAGTATATATATATCTGATGATGATGATGATGATGATGATGATGATGATGAT 1526
DB 177 ATCAACCAACCATATATATATATTTAAATCTTTGTCATTTTCAACATCTTCAAGC 236
QY 1527 GTCTTACCAAGAGTATGATTCATCTCATTTCTGATGATGAAATCTTTGCTCATCTAA 1586
DB 237 ACCTTACCAAGAGAAATTTTATCCCA-----AGAAACATGTTATTTGCTCATTTATA 291
QY 1587 GAAGAAATTCCTCATCTGTTCAAGTTTATCATGATGATGATGATGATGATGATGATGATGAT 1646
DB 292 GAAGCAACTCTTCATCTTGAAGCTTATCATGAGTTT-----GTCAACCTCTT 340
QY 1647 CAGGCTCACTTCACTTTTAAATTCAGTCTCTGCTGCTTTTACCAATCTGATGCTTC 1706
DB 341 CAGGCTTCACTTCA-----AATCTAGTCTTCTGCTGCTTCTTACGATCTGTAGTTAC 395
QY 1707 TTCTTCATGAGTCTTGAACCTCTCAAGTATCATCATGAGGTTGGAATGCACTTCTT 1766
DB 396 TTCTTCATCTG-AGTCTTGAACCTTCAAGTATGATGATGATGATGATGATGATGATGATGAT 454
QY 1767 CCAATTCCTGTTAATATTTATTTTGAAC--TCCATGATATCATGAAATGTTCTTAAT 1823
DB 455 CCAAACTCGTTTATGTTATGTTATTTGACACATTTTATGAAATGCGAAGTCTTACC 514
QY 1824 GGCACCTGGAATGATGATCTTTCCAAAGTTTGAATTTACTTACTTGCAGATTCATC 1883
DB 515 GACATCTTGAATATGATGATCTTTCAGAAAGTTTTCATCTTCTTGTTCATGATCAT- 573

```

OY 1884 CATCCAGAGAAATCCACTTTCATATGCAGTAAAGCTTATGGAATAGTATCTTCAATPA 1943  
 Db 574 ----CAGAGAATCACTGTCTATGGAACGTGTAGCCTTACCTAAAGTATTTCTTAAATPA 629  
 OY 1944 TAAGGCTGAAGTAAGTAATTACTGCTTGATCCATATTTCTG 1985  
 Db 630 TAAATGTGAATATGAAATTACTCTTGATCCATAGGGCTG 671

RESULT 37					
BM978279					
LOCUS	BM978279	690 bp	mRNA	linear	EST 21-FEB-2003
DEFINITION	UI-CF-BC1-aeb-n-17-0-UI.s1 UI-CF-BC1 Homo sapiens cDNA clone				
DESCRIPTION	UI-CF-BC1-aeb-n-17-0-UI 3', mRNA sequence.				
ACCESSION	U000000				

ACCESSION	BM518219	GI:19597547
VERSION	BM518219.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens	
ORGANISM	Homo sapiens	(human)

REFERENCE 1 (bases 1 to 690)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
PUBMED 8908470

Genome Res. 6 (9), 791-806 (1996)  
8889548  
Contact: McCray, PB

University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@iowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics ([www.resgen.com](http://www.resgen.com)) or from Open Biosystems  
([www.openbiosystems.com](http://www.openbiosystems.com)) .  
The following repetitive elements were found in this cDNA  
sequence: 20-690, >TIGER1#DNA/MER2\_type (matched complement)  
Seq primer: M13 FORWARD  
POLYA=yes.

```

FEATURES
source      location/Qualifiers
1. .690

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EC1-aab-n-17-0-UI"
/tissue_type="Lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EC1"
/node="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AAGGCTTAC.
TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383

```

ORIGIN

Query Match	15.5%	Score 310.6	DB 3	Length 690
Best Local Similarity	73.5%	Pred. No. 3	5e-39	
Matches 502	Conservative 0	Mismatches 149	Indels 32	Gaps 7

QY	1064	GAGTCACACAAAGCTTTTGGCTTCCCGACGTCACATATAAAGCTTTGGCTTAATACAGTCTG	1123
Db	20	GAGTCACACAAATTTTTTAATTTTCCAGTACCTATATAAAGTTAATGTTAATCTATATAT	79
QY	1124	AGTCGTAAAGTGTGCATAAGTGTATGTCTAAAAAAACAATACCTTAATTTTAAATG	1183
Db	80	AGCTATTAAGTTTGCATAGCATTAATATCTGCATAATGTACATACCTTAATTTAAAAATA	139
QY	1184	CTTTATTACTAAAAAATGCTAACATCATTTTGAGCATTCAGTGAAGTTGTAATCTTTTGC	1243
Db	140	TTTATATTGCTAAAAAGATGCCTAACGATATATGATGATCTTCAGTAGT-----CTTTGTGTC	194
QY	1244	TGGTGAAGGCTTTTCTTATGTATGACGTAT-----CGGGGGTCAGGTGCTGA	1292
Db	195	TGGAAAGCAGCTCTTGCTCACTGTGTAGTGGCTGTGATGATGCGGTGGTGTGTGTA	254
QY	1293	AGCTTAAGGTGGCTGTGGCAGTTTCTTAA-----ACAACGTGAAGATGCAATATCAG	1347
Db	255	AGGTATGTGGCTCTTGAGATTTCTTTAAATGAGACACAGAAAGCTGCAACATCA	314
QY	1348	TTGACTCTTCTTTTCATGAAAGATTCTCTTAAGTGTGATGCTTTTGATGACATTT	1407
Db	315	CTGACTCTTCTCTTTC--CGAAAGATTATCTGTAGCATGTGTGCTGTGTGATGACCTTT	373
QY	1408	ATGACACAGTGAACCTTTTGAAAAATTGGA--TCAATCTCTCAACCCCTGCTCTGTTA	1466
Db	374	ACCCACAGTGGAACTTTCCAAATAATGGAAGTCAATCTCTCAACCTGCTGCTGTTTT	433
QY	1467	ACAACCTAATTTAATAATATCTGATGACATTTGTCATTTCAAACTTTTCACAGT	1526
Db	434	ATCACTAATGTTATGTAATATCTTAAGTCTTTGTGTATTTCAACAAATGTGTACAGS	493
QY	1527	GCTTTCACAGAGATGATTCATCTCATTTCTCTGATGAGATCTTTGGCTCATTCATTA	1586
Db	494	ATCTTACCGAGAGTAGTTTCCATCTCAGAAACCACTT---TTGGGTCTTAATATTA	549
QY	1587	GAAGAAATTCCTCATCTGTTCAAGTTTATCATGATATGACAGCAATACAGTCATGCTT	1646
Db	550	GAAGCAACTCTCATTTGTTCAAGTTTATCATGATATGACAGCAATTCAGTCACATATG	609
QY	1647	CAGGCTCACTTCACTTTTAATTCAGATTCTCTTGCTGTTTCTACACACATCTGGTTC	1706
Db	610	CAGG-----TTCCACTTCTAAGCTGTAGTTATCTTGTATTTCCACACACATCTGACGTTAC	664
QY	1707	TTGCTTCATTAAGTCTGTGAACC	1729
Db	665	TTGCTTCATTAAGTCTGTGAACC	687

RESULT	38
BP280947/c	
LOCUS	BP280947
DEFINITION	KR080947 Sugano cDNA library, KG-1-C Homo sapiens CDNA clone
ACCESSION	BP280947
VERSION	BP280947.1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 537) Suzuki, Y., Yamashita, R., Shiroya, M., Sakakibara, Y., Chiba, J., Mizushima-Sugano, J., Nakai, K. and Sugano, S.
AUTHORS	

TITLE	Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions
JOURNAL	Genome Res. 14 (9), 1711-1718 (2004)
PUBMED	15342556
COMMENT	Contact: Yutaka Suzuki Department of Virology Institute of Medical Science, University of Tokyo 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan Email: yusuzuki@ims.u-tokyo.ac.jp. Location/Qualifiers
FEATURES	1..537
source	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="KGR09271" /cell_line="KG-1-C" /clone_id="Sugano CDNA library, KG-1-C" /note="g11oma"
ORIGIN	
Query Match	15.5%, Score 310.4; DB 3; Length 537;
Best Local Similarity	80.2%; Pred. No. 3.9e-39;
Matches 417; Conservative	0; Mismatches 91; Indels 12; Gaps 4;
QY	1319 TAAACACACATGTAAGATTGCAAAATACAGTGCCTCTTCCTTATGAAAGATTCTCTC 1378
Db	511 TAAAGCATCGGTGAAGCTTTTGGCATGGTGAAGCTCTTTCATGAAAGATTCTCTG 452
QY	1379 TAGTGTGATGAGCTTTTGGATGATGACATTTTATGACAGTAGAAGCTCTTTGAAAATTGGA- 1437
Db	451 TAGCATGGAGATGCTGTTGATGATGATTTGGCCACAGTAGAAGCTCTTCAAAATTGGAG 392
QY	1438 TCAATCTCTCAAAACCCGCTCTGCTTTACAAACCTAAGTTAATTAATATTGCAATCC 1497
Db	391 TCAATCTCTCAAAACCCGCTCTGCTTTATCACTAAGTTAATTAATATTCTAAATCC 332
QY	1498 ATTGTTGATCTTCAACAATTTTTCACAGTCTTACACAGAGTAGATTCATCTCATTT 1557
Db	331 TTGTTGTCATCTTCAACAATGTTTCCACAGATCTTTACACAGGTAGATTCATCTCADA 272
QY	1558 CCTGAGATGGAATCTTTGCTCATCATTAAGAAGAAATTCATCTCATCTTCAAGTTTATC 1617
Db	271 AACCACTT---TCTTTGGCATCTATAAGAAACAATCTCATCTGTAAAGTTTATC 216
QY	1618 ATGAGATTGCACGAATACAGTCATGCTTTTCAGGCTCATCTTCACTTTTAATTCAGTTCT 1677
Db	215 ATGAGATTGCACCAATTCAGTTACATCTTCAAG-----CTTACTTCAATTCAGTTCT 161
QY	1678 CTGTGCTGTTTCAACCAATCTGTGGTCTTCTCTCATTTGAAGTCTTAACTCTCCAG 1737
Db	160 TTGTCATATTTCCACCAATCTGCACTTACTTCTTCCACAGAGTCTTGAACCCCTCAG 101
QY	1738 TCATCATGAGGGTGGATGATGACTTCTTCCAAATTCCTGTAATATTAATTATTGACC 1797
Db	100 TCATCAATAGGGGTGGAAATCAACTTCTTCCAACTCTGTTAAAGTTGACAAATTGACC 41
QY	1798 TCCCATGATCATGA--ATGTTCTTAATGGCACTGGAAT 1835
Db	40 TCCGCCCATGATCAAGAGTTTATTAATGGCACTTAAAT 1
RESULT 39	
BZ771573/c	1011 bp DNA linear GSS 13-MAR-2003
LOCUS	mcl78a04.g6 HROSMD004 Homo sapiens genomic, genomic survey
DEFINITION	sequence.
ACCESSION	BZ771573
VERSION	BZ771573.1 GI:28945257
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1011)	Hominidae; Homo.			
	Cook, L., Delehaney, K., Fewell, G., Fulton, L., Magrini, V.,			
	Mardis, E., Miner, T., Nash, W., Williams, D. and Wilson, R. K.			
	Homo sapiens Fosmid End Reads			
	Unpublished (2003)			
	Contact: Richard K. Wilson			
	Genome Sequencing Center			
	Washington University School of Medicine			
	Email: submissions@wustl.edu			
	Plate: mct78	row: a	column: 04	
	Class: fosmid ends			
	High quality sequence start: 102			
	High quality sequence stop: 526.			
	Location/Qualifiers			
	1..1011			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
	/clone_id="HPOSMID004"			
	/note="Vector: pcc01fos; Site 1: Eco72i; Human whole			
	genome fosmid library was prepared at Washington			
	University Genome Sequencing Center. DNA was sheared for			
	blunt-ended ligation into pcc01fos inducible vector. DNA			
	was ordered from Coriell Cell Repository's DNA			
	polymorphism discovery resource."			
ORIGIN				
Query Match	15.3%;	Score 306.6;	DB 9;	Length 1011;
Best Local Similarity	77.0%;	Pred. No. 1.4e-38;		
Matches 441;	Conservative 0;	Mismatches 119;	Indels 13;	Gaps 5;
QY	1319	TAAACAACAGTGAAGATTGCAATATGATGATGATCTCTCTTTCATGTAAGATTCTCTC	1378	
DB	677	TAAACAACAGTGAAGATTGCAAGCCACATGATGATGATCTCTCTTTCATGTAAGATTCTCTC	618	
QY	1379	TAGTGTGATGATCTCTTTCATGATGATTTTA--TGCACATGAACTCTTGAAGATTGG	1436	
DB	617	TAGCTATGATGATCTCTTTCATGATGATTTTACCTACACAGTAAATGCTTCAAAATGTC	558	
QY	1437	A-TCAATCTCTCAAACTCTCTCTCTTTCATGAACTTAATTAATTAATCTGAT	1495	
DB	557	AGTCAGTCTCTCAAACTCTCTCTCTCTTTCATGAACTTAATTAATTAATCTGAT	498	
QY	1496	CCATGTTGTCATTTCAACAATTTTCACAGTCTCTCAAGAGTAGATTCATCTCAT	1555	
DB	497	CCATTATGCTGTTTCACAATGTTTCACAGATTTTCACAAGATGTTTCCTGATCT	438	
QY	1556	TTCCTGAGATGGA--ATCTTGTCTCATCATTAAGAAAGAAATTCCTCATCTGTTCAAGTTT	1611	
DB	437	TTCAGGAACACATCTTCTTCTCATCAACAAGAGAGGCTTCATCTGTTGAAGTT	378	
QY	1614	TATCATGAGATTGACGACATACAGTCATGCTTTCAGGCTCTCATCTTTCATTTATTCAG	1677	
DB	377	GGTCAATGAGATTTCAGTAATTCAGTCATCTTTCAGGCTCTCTCTC-----TAATTCAG	323	
QY	1674	TTCTCTGCTGTTTTCACACATCTGCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	1733	
DB	322	TTCTCTGCTGTTTTCACACATCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	263	
QY	1734	CAAGTCATCAGAGGAGGTTGGAATGAGATCTTTCACAAATTCCTGTTAATTTATTTT	1793	
DB	262	CAAGTCATCTGAGAGGAGGTTGCAATCAGCTTTTCCAAATCTCTGTTAATTTGATTTTG	203	
QY	1794	GA---CTTCCATGATCATGATGTTCTTAAATGCACTGGAATGATGATCTTTCACA	1855	
DB	202	GACCTCTCTCCATGATCATCAATCTTCTTAATGCACTGAAATGATTAATCTTTCACA	1433	
QY	1851	AAAGTTTCAATTACTTATGATGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	1883	
DB	142	GAAAGTTTCAATATCTCTGCTCAAGATCATCT	110	

CC765346 723 bp DNA linear GSS 27-JUN-2003  
CH240\_57C13.TV CHORI-240 Bos taurus genomic clone CH240\_57C13,  
genomic survey sequence.  
CC765346  
CC765346.1 GI:32311849  
GSS.  
Bos taurus (cow)  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 723)  
Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,  
Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,  
Mowack,J.E., de Jong,P.J. and Lewin,H.A.  
Bovine BAC end sequences from CHORI-240 library  
Unpublished (2003)  
Other\_GSSs: CH240\_57C13.TU  
Contact: Harris Lewin  
Department of Animal Sciences  
University of Illinois at Urbana Champaign  
1201 W. Gregory Dr., Urbana, IL 61801, USA  
Tel: 217 333 5598  
Fax: 217 244 5617  
Email: h-lewin@uiuc.edu  
Clones are derived from the bovine BAC library CHORI-240  
(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library  
availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
([http://www.chori.org/bacpac/ordering\\_information.htm](http://www.chori.org/bacpac/ordering_information.htm)). This work  
was undertaken as part of the International Bovine BAC Mapping  
Consortium (IBMC) by University of Illinois at Urbana  
Champaign, USA with funds provided by grant No. AG202-34480-11828  
from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock  
Genome Sequencing Initiative)  
Plate: 57 row: C column: 13  
Seq primer: T7  
Class: BAC ends.

Query Match	15.3%;	Score 306;	DB 9;	Length 723;
Best Local Similarity	73.9%;	Pred. No. 1.8e-38;		
Matches 509;	Conservative	0;	Mismatches 138;	Indels 42;
				Gaps 8;

QY 1313 TAAACAACAAGTGAAGATTGCACAAATGACGTGACCTCTTCCTTTCATGAAAGATTCTCTC 1378  
Db 19 TAAACAACAACANNAGATTGCG-ATGTCGATGATGATCTCCCTTCATGATGATTCTCTG 77  
QY 1379 TAGGTGTGATGCTTTTGTGATAGCATTTTATGACAGTAGAAACTCTTTGAAATTTGCA- 1433  
Db 78 TAGGTGTGATGACTGTTGTGTAGCATTTTACTTACAGTAGAACTCTTTCAAAATTTGAG 137  
QY 1438 TCATTCCTTCGAAACCCGCTCTGCTTTAAACAACCTAGTTAAATATATTTGCAATCC 1497  
Db 138 TCAGTCCCTTAAACCCCGC-----TGCCTAATGTCTTAAATCT 177  
QY 1498 ATTGTGTCATTTAAACAATTTTCAACAGTGTCTTACACAGAGAGATGATTTCCATCTCATTT 155

RESULT	41
CEO27248/c	
LOCUS	CEO27248
DEFINITION	746 bp DNA linear GSS 24-SEP-2003 t1gr-gss-dog-17000344638788 Dog Library Canis familiaris genomic,
ACCESSION	CEO27248
VERSION	CEO27248.1 GI:35045787
KEYWORDS	GSS.
SOURCE	Canis familiaris (dog)
ORGANISM	Canis familiaris Bukaryotic; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE	1 (bases 1 to 746) Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C. The dog genome: survey sequencing and comparative analysis Science 301 (5641), 1898-1903 (2003)
TITLE	
JOURNAL	

COMMENT	Contact: Kirkness EF The Institute for Genomic Research Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-838-0280 Fax: 301-838-0208 Email: ekirkness@tigr.org Class: Shotgun.
FEATURES	Location/Qualifiers
source	1..746 /organism="Canis familiaris" /mol_type="genomic DNA" /strain="Standard poodle" /db_xref="taxon:9615" /clone_lib="Dog Library" /note="Site 1: Bacti; Libraries were prepared from peripheral blood"
ORIGIN	

Query Match 15.3%; Score 305.6; DB 9; Length 746;  
 Best Local Similarity 72.2%; Pred. No. 2,1e-38;  
 Matches 520; Conservative 0; Mismatches 164; Indels 36; Gaps 8;

1287 TCGTGAAGCTTAAAGGTGGCTGTGGCAAGTTCTTAAACG-----ACAGTGAAGATTGCA 1340  
 |||||  
 740 TCGTGAAGGTGGGTGGCTGTGGCAAGTTCTTAAATAGGACAGTGAAGTTGTTGGC 681  
 |||||  
 1341 AATATGAGTGAATCTTCTTATGAAAGATTCTCTAGTGTGATGCTTTTGATA 1400  
 |||||  
 680 AATATGAGTGAATCTTCTTATGAAAGATTCTCTAGTGTGATGCTTTTGATA 621  
 |||||  
 1401 GCATTTTATGACAGTGAAGATTCTTGAAGAAATTTGATCAATCTTCAACCTGCTCT 1460  
 |||||  
 620 GCATTTTATGACAGTGAAGATTCTTGAAGAAATTTGATCAATCTTCAACCTGCTCT 573  
 |||||  
 1461 GCTTTAACAACCTAAGTAAATATATATATATGATTCATCTGTTGTCATTTCAACATTTT 1520  
 |||||  
 572 TCGTTTATCAACAGGTGTATGTAATATTTCTTAATCTTGTGTCATTTCAATAGTCTCT 513  
 |||||  
 1521 CACAGTGTCTTACCGAGAGTATGATTCATCTGATTTCTCGATGAGTGAATCTTGTCTCAT 1580  
 |||||  
 512 CACCGGTGTTTACCAAGAGTATTCATCTCGAGAAACCACTT-----TTCTGATCAT 457  
 |||||  
 1581 CCAATGAAGAAGAAATCTCTATCTGTTCAAGTTTATCATGAGATTCAGCAATACAGTCA 1640  
 |||||  
 456 GTGTAAAGAGTAAATTTCTTCACTCTTCAAGTTTATCATGAGATTCAGCAATACAGTCA 397  
 |||||  
 1641 TGTCTTCAAGGCTCACTTCACTTAAATTCAGTCTCTGCTGTTTCAACATCTGT 1700  
 |||||  
 396 CTCTCTTCAAGGCTCACTTCACTTAAATTCAGTCTCTGCTGTTTCAACATCTGT 341  
 |||||  
 1701 GGTCTCTCTCTCATGAGTGAATCTGTAACCTCCATGATCATGAGGTTGGAATTCGA 1760  
 |||||  
 340 AGGTACTTCTCTCATGAGTGAATCTGTAACCTCCATGATCATGAGGTTGGAATTCGA 282  
 |||||  
 1761 CTCTCTTCAAGGCTCACTTCACTTAAATTCAGTCTCTGCTGTTTCAACATCTGT 1820  
 |||||  
 281 CTCTCTTCAAGGCTCACTTCACTTAAATTCAGTCTCTGCTGTTTCAACATCTGT 225  
 |||||  
 1821 AATGAGCACTGTAAGTGAATCTTCCAAAGGTTTCAATTTAGTCAAGTCA 1880  
 |||||  
 224 AATGAGCACTGTAAGTGAATCTTCCAAAGGTTTATTTACTTCTTCAAGTCA 166  
 |||||  
 1881 ATTCATCCAGAGATCACTTCAATGCCAGTTATAGCTTATGAGATGATTTCTTCA 1940  
 |||||  
 165 AT-----CAGAGGCATCACTATCTATGAGGCTCTGAGCTTATTAATGATTTCTTAA 111  
 |||||  
 1941 TAATTAAGGCTTGAAGTGAATTTACTCTTGAATCCATTTTCTGCAAAATAGATGTTGTG 2000  
 |||||  
 110 CAATTAAGGCTTGAAGTGAATTTACTCTTGAATCCATTTTCTGCAAAATAGATGTTGTG 51

RESULT 42  
 BP981537 674 bp mRNA linear EST 23-JAN-2001  
 LOCUS BP981537/c 602309408F1 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:4400857 5',  
 DEFINITION mRNA sequence.

ACCESSION BP981537  
 VERSION BP981537.1 GI:12384349  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: LNLN10106 row: 0 column: 02  
 High quality sequence stop: 664.  
 Location/Qualifiers

## FEATURES

1..674  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4400857"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 88"  
 /note="Organ: small\_intestine; Vector: pCMV-SPORT6;  
 Site\_1: Not1; Site\_2: SalI; Cloned unidirectionally;  
 oligo-dr primed. Average insert size 1.767 kb. Library  
 enriched for full-length clones and constructed by Life  
 Technologies. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 15.3%; Score 305.4; DB 2; Length 674;  
 Best Local Similarity 73.0%; Pred. No. 2,3e-38;  
 Matches 465; Conservative 0; Mismatches 156; Indels 16; Gaps 5;

1368 AGATTTCTCTAGTGTGATGATCTTTTGAATGAGATTTTATGCAAGTGAAGCTTCTT 1427  
 |||||  
 660 AATATTTCTGTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 601  
 |||||  
 1428 GAATATGATCAATCTCTCAACCTGCTGCTTTTAACTTAATTAATTA 1487  
 |||||  
 600 CAATATGAGATCAATCTCTCAACCTGCTGCTTTTAACTTAATTAATTA 543  
 |||||  
 1488 TTCTGAATCAATGTTGTCATTTCAACATTTTCAAGTGTCTTCAACAGAGATGATTC 1547  
 |||||  
 542 TTCTGAATCAATGTTGTCATTTCAACATTTTCAAGTGTCTTCAACAGAGATGATTC 483  
 |||||  
 1548 CATCTCATTTCTGAGATGGAATCTTGTCTCATTCATAGAGAAATTCCTATCTGTTC 1607  
 |||||  
 482 CATCTCAAGAAACCACTT-----TCTTGTCTCATTCAGAGAAACCACTTCTATCTGTTC 427  
 |||||  
 1608 AAGTTTATCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1667  
 |||||  
 426 AAGTTTATCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 373  
 |||||  
 1668 TTCCAGTCTCTGCTGTTTCAACATCTGCTGCTTCTCTCATGAAAGTCTTGA 1727  
 |||||  
 372 TTCTAGTCTCTGCTGTTTCAACATCTGCTGCTTCTCTCATGAAAGTCTTGA 313  
 |||||  
 1728 CTTCTCAAGTCAATCAATGAGGTTGGAATTCATCTTCCAAATCTGTTAATTA 1787  
 |||||  
 312 TCCCTCAAGTCAATCAATGAGGTTGGAATTCATCTTCCAAATCTGTTAATTA 253  
 |||||  
 1788 TATTTTGAAC--CTCCATGAATCATGAATGTTCTTAATGAGCACTGGAATGAGTGAATC 1844  
 |||||  
 252 TATTTTGAACCTTCTCCATGAATCACAAAATGTTCTTAATGAGCACTGGAATGAGTGAATC 193  
 |||||  
 1845 TTTCCAAAGTTTTCAATTTACTTATGTCAGATTCATTCATTCAGAGATTCATTTCA 1904  
 |||||  
 192 TTTCCAGAGGTTTCAATTTACTTGGCAAAATTCATCAAAATTAATTAATTAATTAATTA 193  
 |||||  
 1905 ATGCCAG-TTATAGCCTTATGAGATGATTTCTTCAATTAATAGGCTTGAAGTGAAT 1963  
 |||||  
 132 AGCATAGCCTCAAAATGATTAATTAATTTCTCAAAATTAATAGATGAGTGAAGTGAAT 73  
 |||||  
 1964 TACTCTGATTCATTTTCTGCAAAATTAATGATGTTGTG 2000  
 |||||  
 72 TACTCTGATTCATGAGGCTGAGAAATGTTGTG 36

RESULT 43  
 BU680222  
 LOCUS  
 DEFINITION  
 UI-CF-DUI-abe-n-08-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone  
 BU680222  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 696)  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 8889548  
 CONTACT: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul.mccray@uiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com) or from Open Biosystems  
 (www.openbiosystems.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 18-687, >TIGER1#DNA/MER2\_type (matched complement)  
 Seq primer: M13 FORWARD  
 POLY-A=yes

FEATURES  
 source  
 1. 696  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-DUI-abe-n-08-0-UI"  
 /tissue\_type="Primary Lung Epithelial Cells"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_1ib="UI-CF-DUI"  
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
 UI-CF-DUI is a normalized cDNA library containing the  
 following tissue(s): Primary Lung Epithelial Cells The  
 library was constructed according to Bonaldo, Lennon and  
 Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an EcoR I adaptor, digested with Not I, and cloned  
 directionally into pT73-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dT)18 tail. The sequence tag for this  
 library is GGCTGTAGGC.  
 TAG\_TISSUE=Lung Epithelial Cells Tissue nos 359-368  
 TAG\_LIB=UI-CF-DUI  
 TAG\_SEO=GGCTGTAGGC"

ORIGIN  
 Query Match 15.2%; Score 304.2; DB 5; Length 696;  
 Best Local Similarity 73.7%; Pred. No. 3.5e-38;  
 Matches 505; Conservative 0; Mismatches 145; Indels 35; Gaps 8;  
 1166 TACCTTAATTTAAATATTTATGCTAAAGATGCTAATCTAAGTCTTCAGT 1225

Db 18 TACCTTAATTTAAATATTTATGCTAAAGATGCTAATCTAAGTCTTCAGT 77  
 Qy 1226 GAGTTGTAATCTTTTCTGCTGCTGAGAGCTTTTCTTATGATGACGATCGGGGGCA- 1284  
 Db 78 GAGTCT-----TTGTTGCTGGAAGACAGTCTTCTCCTACGTTGAAGCTGCTGAAGAT 132  
 Qy 1285 -----GGTGTGAAGCTTAGGGTGGCTGTGACAGTTCTTAA-----ACAAAG 1329  
 Db 133 CGGT 132  
 Qy 1330 TGAAGATTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1389  
 Db 193 GGAAGGTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251  
 Qy 1390 GCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1448  
 Db 252 GCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311  
 Qy 1449 AAACCGCTCTGCTTTTAAACCTTAATATATATATATATATATATATATATATATATAT 1508  
 Db 312 AAACCTCTGCTGCTTTTAAACCTTAATATATATATATATATATATATATATATATATAT 371  
 Qy 1509 TTCACAAATTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1568  
 Db 372 TTCACAAATTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 428  
 Qy 1569 ATCTTGTCTTCATTCATTAAGAAATTTCTCATCTGTTCAAGTTTATCATGAGATTGCA 1628  
 Db 429 -TTGTGCTCTTTTATTTAGAGACACCTTCATTTGTCAGTTTATCATGAGATTGCA 487  
 Qy 1629 GCAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1688  
 Db 488 GCAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 542  
 Qy 1689 TACCAATCTGTTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1748  
 Db 543 CACCAATCTGTTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 602  
 Qy 1749 GGTGGAATGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1805  
 Db 603 GGTGGAATGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 662  
 Qy 1806 ATCATGAATGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1830  
 Db 663 ATCATGAATGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 687

RESULT 44  
 DN100706  
 LOCUS  
 DEFINITION  
 1000083 MARC 4P1G Sus scrofa cDNA 3', mRNA sequence.  
 DN100706  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Sus scrofa (pig)  
 Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suidae; Suidae;  
 Sus.  
 1 (bases 1 to 849)  
 Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,  
 Noneman,D.J., Wray,J.E. and Keeler,J.W.  
 Porcine EST collection using a normalized library constructed from  
 embryos representing early developmental stages  
 Unpublished (2003)  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called with phred v0.020425.c and  
 trimmed with the aid of the trim\_alt option. Vector identified with



cross\_match v0.990329.  
 Plates: TMW8069 row: P column: 22  
 Seq primer: TGAAGGCGACACTCGAGG.  
 Location/Qualifiers

# FEATURES

source

1..849  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /issue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_1ib="MARC 4P1G"  
 /note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
 library made with combined RNA from day-10, day-13,  
 day-15, day-25, and day-30 whole embryos."

## ORIGIN

Query Match 15.2%; Score 304; DB 8; Length 849;

Best Local Similarity 69.1%; Pred. No. 3,6e-38;  
 Matches 612; Conservative 0; Mismatches 215; Indels 59; Gaps 12;

1121 TGTAGTCTGTAAAGTGTGAAGTGTATGTCTAAACCAACATACCTTAATTTAA 1180  
 |||||  
 3 TTGGCTTTTGTCTTTTGAAGGCGCGCCCAATGATACATCTTCAATAAAA 62  
 |||||  
 1181 ATGCTTATTAATAAATGCTAATCAATCATTTGAGCATTCAGTGTGTATCTTT 1240  
 |||||  
 63 ATACTTATTTGCT-AAAATTGCTACATCATCTTAAGCTTTTGAAGTTGATCTTT 121  
 |||||  
 1241 TCTGTGTGAAGTCT- -TTTCTTATGATGATGAT- - -CGGGGTCAAGTGC 1289  
 |||||  
 122 TATGTGTAAAGGCTCTTGGCTTGATATGATGATGCTGGCCAAACAGGGTGGCTGT 181  
 |||||  
 1290 TGAAGCTTAAAGGCTGTGGCTGTGGCTTTCTTAA-ACAACAGTGAAGATTGCAATTCAG 1347  
 |||||  
 182 TGAAGCTGTGAAGGCTTATGAGCAATTTCTTAAATAGAAACAATGAATACCAATGCG 241  
 |||||  
 1348 TTGACTCTTCTTCAATGAAGATTTCTCTAGTGTGATGCTTTTGTATGATGAT 1407  
 |||||  
 242 TTGGCTACTCTTCAATGAAGATTTCTCTGTGATGATGATGATGATGATGATGAT 301  
 |||||  
 1408 ATGCAAGTGAAGATTTCTTGAAGATTTGATCAATCTCTCAAAACCTGTCTGCTTAA 1467  
 |||||  
 302 CCACAGCAGCAATTTTTCAAAATAGAGTCACTTCTTAATCTGCACTGCTTCA 361  
 |||||  
 1468 CAACCTTAATTAATTAATTTCTGAATTCATTTGCTATTTCAACATTTTCAACAGT 1527  
 |||||  
 362 CAAC- -TTCAATTAATTAATTTCTGAATTTCTTGTGATTTTCAACATTTTCA- - 412  
 |||||  
 1528 TCTTCAACAGGATGATTCATCTCATTTCTCGAGATGAGATCTTGTCTCATCATTAAG 1587  
 |||||  
 413 - - - - -CAATAGGCTTATCTTCAAGAAACATTT- - -TCTTGTCTTATACATTAAG 458  
 |||||  
 1588 AAGAAATTCCTCATCTGTCAAGTTTATCATGATGATGAGCAATACAGTCAATGCTTTC 1647  
 |||||  
 459 AAGCAATTCCTTAATTCATTAAGTTTATCATGATGATGAGCAATACAGTCAATTAATTC 518  
 |||||  
 1648 AGGCTTCACTTCACTTAAATTCAGTTCTTGTCTGTTTCTTACCAATCTGTGCTCT 1707  
 |||||  
 519 ATG- - - -TTTCAATTTCTAATTCGATTTCCCTGCTTCTTCTTCAATCTGCAATCACT 573  
 |||||  
 1708 TCTTCAATTAAGTCTTGAAGCTCTTCAAGTCAATCAAGGCTTGAATGCAATCTTCTTC 1767  
 |||||  
 574 TACTTCAGTGAAGCTTGAAGCCTT- - - - -TCTGTGAGGCTTACCAACCAATTTTTC 625  
 |||||  
 1768 CAATTCCTGTAAATTAATTAATTTTGAAC- - -TCCATGATCATGATGATGCTTATAG 1824  
 |||||  
 626 CAATTCCTTAATTAATTAATTTTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 685  
 |||||  
 1825 GCACTTGAAGTGAATGCTTTCGCAAAAGTCTTCAATTTCAATTTAGTCCAGATCC 1884  
 |||||  
 686 ACATCTGAATGTAATTTCTTCCCAAAAGTCTTCAAAACCAATTCGCCGAGATCCAT- - 743  
 |||||  
 1885 ATCCAGAGATCCACTTTCATGCGAGTATAGCTTATGGAATGATTTCTTCAATTAAT 1944

Db 744 - - - - -CAGAGAAATCACTTTCTAATGCAAGTCACTGCTT-TCAAATGATTTCTTCAATAA 799  
 |||||  
 QY 1945 AAGGCTTGAAGTGAATTAATCTCTTGAATCAATTTCTGCAAAAT 1990  
 |||||  
 Db 800 AAGATCTAATAATTAATGATGACTCTTATCTTCAATCTGAGATCTGAGAT 845  
 |||||

## RESULT 45

CA308451/c 620 bp mRNA linear EST 05-AUG-2004

LOCUS UI-H-F11-bhz-b-08-0-UI.s1 NCI CGAP\_F11 Homo sapiens cDNA clone

DEFINITION UI-H-F11-bhz-b-08-0-UI 3', mRNA sequence.

ACCESSION CA308451

VERSION CA308451.1 GI:24471505

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Homo.

1 (bases 1 to 620)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Gary W. Hunninghake, U of I

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained

from Dr. M. Bento Soares, bento-soares@uiowa.edu

The following repetitive elements were found in this cDNA

sequence: 18-620, >TTGGER1#DPM/MER2\_type

Seq primer: M13 FORWARD

POLYA=Yes.

## FEATURES

source

Location/Qualifiers

1..620

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-H-F11-bhz-b-08-0-UI"

/issue\_type="Alveolar Macrophage"

/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies)"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

NCI CGAP F11 is a normalized cDNA library constructed from

a pool of 81 RNA samples from Alveolar Macrophages

challenged with different treatments. The mRNA samples

were a mixture of these conditions (times refer to

incubations following isolation by bronchoalveolar lavage)

(some normal donor macrophages were cultured in some of

the conditions, other donor macrophages in different

conditions). The mRNA samples were pooled for library

construction. Control 0 hours; control 3 hours; control 24

hours; LPS 100 ng/mL, 3 hours; LPS 100 ng/mL, 24 hours;

PMA 10 ng/mL, 3 hours; PMA 10 ng/mL, 24 hours; Klebsiella

moi 10, 3 hours; Klebsiella moi 10, 24 hours; Staph aureus

moi 10, 3 hours; Staph aureus moi 10, 24 hours; Adenoviral

vector (Ads CMV egfp), moi 500, 3 hours; Adenoviral vector

(Ads CMV egfp), moi 500, 24 hours; wt adenovirus moi 500,

3 hours; wt adenovirus moi 500, 24 hours; Ad vector + LPS 3

hours; wt adenovirus + LPS 24 hours; wt adenovirus + LPS 3

hours; wt adenovirus + LPS 24 hours; The library was

normalized according to Bonaldo, Lennon and Soares, Genome

Research, 6:791-806, 1996. First strand cDNA synthesis was

primed with an oligo-dT primer containing a Not I site.

Double stranded cDNA was ligated to an EcoR I adaptor,

digested with Not I, and cloned directionally into



p1773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)<sub>18</sub> tail. The sequence tag for this library is GGGCATGCGG. The tissue was provided by Dr. Gary W. Hummingsake of the University of Iowa.  
TAG TISSUE=Human Lung Alveolar Macrophage  
TAG LIB=UI-H-FT1  
TAG\_SEQ=GGCCATGCGG"

# ORIGIN

Query Match 15.2%; Score 303; DB 6; Length 620;  
Best Local Similarity 73.4%; Pred. No. 5.6e-38;  
Matches 467; Conservative 0; Mismatches 136; Indels 33; Gaps 5;

QY	1330	TGAAGATTGCAATATCAGTTGACTCTTCCTTATGAAAGATTCTCTAGTGTGAT	1389
DB	620	TGAAGCTTGCCACATCAATGACCTTCTCTTCAAGAAAGATTCTCTGTAAGCATGTG	561
QY	1390	GCTTTTGAATGATTTTATGACAGTAGAACTTTTGAAGATGATGATCTCTCA	1449
DB	560	ATGATGTGC-----ATTTACACACAGAACTTCTTCAAAATTGAGTCAATCTCTCA	507
QY	1450	AACCTGCTCTGCTTTTAAACAACCTTAATTAATTAATTCGATCATTTGTCATT	1509
DB	506	AACCTGCTGCTCTTTATTTGACTTAATTAATTAATTAATTCGATCATTTGTCATT	447
QY	1510	TCAACAATTTTCAAGTGTCTTCAACAAGATGATTCATCTCAATTCCTGAGATGAA	1569
DB	446	TCAACAATGTTGACACACTTTCACACAGAGTAGATTCCATCCAGAAATCATT----	391
QY	1570	TCTTGTCTCATCATTAAGAAATCCATCTCATGTTTCAAGTTTATCATGAGATGCGAG	1629
DB	390	TCTTCACTCATTCATTAAGAAAGCAATCTCATGAGAAAGTTTAAATGAGATTGCGAG	331
QY	1630	CAATACATGATGCTTTCAGAGGCTCACTTCACTTTAATTCAGTTCTCTGCTGTTCT	1689
DB	330	CAATTGAG-----GCTCATTCTTAATTTGATTTGCTATTCT	288
QY	1690	ACCACATCTGTGTTCTTCTCTCATTTGAAGTCTTGAACCTCTCAAGTCAATCATGAAG	1749
DB	287	ACCACATGAGGAGTTACTTCTCTCATGTAAGTCTTGAAGTCTTCAAAATCATCCAGAG	228
QY	1750	GTTGAATGACTTCTTCCAAATCTCTGTAATTAATTAATTTGAGCTCCCATGAATCA	1809
DB	227	TTTGAATGAGCTTTTCCAAACTCTTATTAAGTGAATATTTGA-CTTCAATGCATCA	169
QY	1810	TGAATGTTCTTAATGAGCAGTGAATGTAATCTTCCAAAAGGTTTCAATTTACTT	1869
DB	168	TGAATGTTCTTAATGATATCTGAATGTAATCTTCCAAAAGGCTTCAATTTGCTT	109
QY	1870	AGTCCAGATTCATTCATCCAGAGATCCATTTCATGAGCAGTTATAGCCTTATGGAATG	1929
DB	108	TGCCCGAGACCAT-----CAGAGATCAATTAATCTATGAGCAGTATAGCCTTATGGAATT	54
QY	1930	TATTTCTTCAATTAATGAGCTTGAAGTGAATTA	1965
DB	53	TATTTCTTAATTAAGGAGCTTGAAGTCAAGATTA	18

Search completed: January 22, 2006, 21:08:38  
Job time : 5438.83 secs

Page Blank (upside)

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 ; Search time 6805.08 seconds  
(without alignments)  
16706.187 Million cell updates/sec

Title: US-09-728-552A-3\_COPY\_1\_2000

Perfect score: 2000

Sequence: 1 gattctctctgcctcagcct.....ctgcgaatagatgttctg 2000

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5881141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_scs:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_hvg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2000	100.0	40917	6	AX033911 Sequence
2	2000	100.0	80622	8	AF222855 Homo sapi
3	1966	98.3	112442	8	AL355340 Human DNA
4	1966	98.3	176432	14	AL391648 Homo sapi
5	1611.4	80.6	41008	6	AX033912 Sequence
6	1592.8	79.6	69058	8	AF222856 Homo sapi
7	1592.8	79.6	80202	8	AF222854 Homo sapi
8	1575.2	78.8	80155	8	AF042484 Homo sapi
9	636.8	31.8	2418	6	AR612649 Sequence
10	634	31.7	10020	8	HSU49973 Human t19ge
11	526.6	26.3	158276	14	AC092165 Homo sapi
12	526.6	26.3	158276	14	AC012033 Homo sapi
13	526.6	26.3	179816	14	AC074188 Homo sapi
14	526.6	26.3	210301	8	AF307337 Homo sapi
15	523.4	26.2	2641	8	BC035143 Homo sapi
16	522.2	26.1	60169	8	AL596257 Human DNA
17	518.8	25.9	51343	8	AC108017 Homo sapi
18	516.8	25.8	135818	14	AC158447 Loxodonta

19	516.8	25.8	141510	14	AC157635	AC157635 Loxodonta
20	516	25.8	3895	6	AX833395	AX833395 Sequence
21	516	25.8	3895	8	AK095077	AK095077 Homo sapi
22	513.2	25.8	115958	8	AC004736	AC004736 Human chr
23	515.2	25.8	115971	8	AC124056	AC124056 Homo sapi
24	515.2	25.8	145713	8	AC124301	AC124301 Homo sapi
25	514.4	25.7	178098	8	HS168145	AL022069 Human DNA
26	511.2	25.6	73991	8	AL357332	AL357332 Human DNA
27	511.2	25.6	114056	8	AL590674	AL590674 Human DNA
28	511.2	25.6	128479	8	AL596266	AL596266 Human DNA
29	511	25.6	99025	8	HSU45N11	AL132996 Human DNA
30	510.4	25.5	172061	14	AC027156	AL132996 Homo sapi
31	510.4	25.5	143747	8	HS370M22	282206 Human DNA s
32	508.2	25.4	173691	8	AC022101	AC022101 Homo sapi
33	507.6	25.4	98517	8	HS366N23	AL021331 Human DNA
34	506.2	25.3	110907	8	AC084703	AC084703 Homo sapi
35	506.2	25.3	183794	8	AC013752	AC013752 Homo sapi
36	505.2	25.3	156320	8	AC022028	AC022028 Homo sapi
37	505.2	25.3	159771	8	AL139142	AL139142 Human DNA
38	505.2	25.3	162163	14	AC021635	AC021635 Homo sapi
39	505.2	25.3	169424	14	AL929203	AL929203 Homo sapi
40	505.2	25.3	188759	6	CQ870480	CQ870480 Sequence
41	504	25.2	175210	8	CNS07B65	AL590338 Human chr
42	504	25.2	200853	8	CNS01DSQ	AL121839 Human chr
43	503.4	25.2	68048	8	AC004764	AC004764 Homo sapi
44	503.4	25.2	126034	8	AC008670	AC008670 Homo sapi
45	503.4	25.2	178942	8	AL357520	AL357520 Human DNA
46	503.4	25.2	245383	14	AL161653	AL161653 Homo sapi
47	501.8	25.1	136961	8	AC136290	AC136290 Homo sapi
48	501.8	25.1	140549	8	AP006192	AP006192 Homo sapi
49	501.8	25.1	163908	8	AC005411	AC005411 Homo sapi
50	501.8	25.1	200000	8	AP000493	AP000493 Homo sapi
51	501	25.1	179590	8	AC115986	AC115986 Homo sapi
52	501	25.1	180005	14	AC013532	AC013532 Homo sapi
53	501	25.1	180917	8	AC018517	AC018517 Homo sapi
54	500.6	25.1	185802	14	AC073843	AC073843 Homo sapi
55	500.6	25.0	170333	14	AL139019	AL139019 Homo sapi
56	500.6	25.0	192589	8	AC093157	AC093157 Homo sapi
57	500.4	25.0	3376	8	HSW803941	AL832631 Homo sapi
58	500.2	25.0	138350	14	AC010801	AC010801 Homo sapi
59	500	25.0	139273	8	AL596275	AL596275 Human DNA
60	500	25.0	187999	14	AC026928	AC026928 Homo sapi
61	496	24.8	213721	8	HS172B20	AL022319 Human DNA
62	495.2	24.8	211652	8	AL391825	AL391825 Human DNA
63	494.6	24.7	125565	8	AL359762	AL359762 Human DNA
64	493	24.6	179844	8	AC009265	AC009265 Homo sapi
65	492.8	24.6	103488	8	AL805915	AL805915 Human DNA
66	492.8	24.6	153771	14	AC037463	AC037463 Homo sapi
67	492.8	24.6	160396	8	AL592063	AL592063 Human DNA
68	492.8	24.6	168460	14	AC011585	AC011585 Homo sapi
69	492.8	24.6	182912	8	AC068060	AC068060 Homo sapi
70	492.8	24.6	182913	8	AC068556	AC068556 Homo sapi
71	492.8	24.6	201133	14	AL807235	AL807235 Homo sapi
72	492.6	24.6	147760	8	AC011846	AC011846 Homo sapi
73	490.8	24.5	202874	8	AL389888	AL389888 Human DNA
74	490.6	24.5	140100	8	AC016778	AC016778 Homo sapi
75	490	24.5	168153	8	AF127577	AF127577 Homo sapi
76	490	24.5	340000	8	AC012499	AC012499 Homo sapi
77	490	24.5	340000	8	HS21C008	AL163208 Homo sapi
78	489.8	24.5	214175	8	CNS01DRN	AL118557 Human chr
79	489.4	24.5	179902	8	CNS01DRN	AL117672 Human chr
80	489	24.4	195981	8	AC007652	AC007652 Homo sapi
81	488	24.4	127683	8	AC003683	AC003683 Homo sapi
82	486.8	24.3	175579	8	AC007768	AC007768 Homo sapi
83	486.8	24.3	199702	14	AC087178	AC087178 Homo sapi
84	486.4	24.3	137520	4	BCA45965	AJ45965 Equus cab
85	486.4	24.3	188165	8	CNS00003	AL049971 Human chr
86	485.4	24.3	163106	14	AP002375	AP002375 Homo sapi
87	485.4	24.3	169755	14	AP001199	AP001199 Homo sapi
88	485.4	24.3	190066	8	AC010127	AC010127 Homo sapi
89	485.4	24.3	191194	14	AC021673	AC021673 Homo sapi
90	485.2	24.3	194635	8	AC012087	AC012087 Homo sapi
91	484.6	24.2	196799	8	AC146074	AC146074 Pan trogl

C 92	484	24.2	172345	8	AC084859	Homo sapi
C 93	483.8	24.2	147394	8	HS353H6	AC0282577 Human DNA
C 94	483.6	24.2	153364	14	AC0093787	Homo sapr
C 95	483.6	24.2	156645	8	AC112317	Homo sapi
C 96	483.4	24.2	94817	8	HS981123	Human DNA
C 97	483	24.1	138591	14	AL357733	Homo sapr
C 98	483	24.1	138905	14	AC161886	Pan trogl
C 99	482.6	24.1	177472	8	AC008786	Homo sapi
C 100	482.4	24.1	144048	8	AC013762	Homo sapi

## ALIGNMENTS

RESULT 1	LOCUS	AX033911	AX033911	40917 bp	DNA	1 linear	PAT 21-SEP-2000
DEFINITION	Sequence 3 from Patent WO9851790.	AX033911					
ACCESSION	AX033911.1	GI:10280479					
VERSION							

SOURCE	unidentified
ORGANISM	unidentified
	unclassified sequences.

AUTHORS Cancellia, M.R., Choo, K.H. and Du, S.D.  
TITLE A novel nucleic acid molecule  
JOURNAL Patent: WO 9851790-A 3 19-NOV-1998;

FEATURES  
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART  
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)  
Location/Qualifiers

```

/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

```

## ORIGIN

Query Match	100.0%;	Score	2000;	DB	6;	Length	40917;
Best Local Similarity	100.0%;	Pred. No.	1.7e-300;				
Matches 2000; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0

QY	1	GAATCTCGGTCTAGAGCTCCCAAGTACTAGAGTTACAGGTGCAGACCAACGTCCA	60
Db	1	GAATTCCTGCTCAGAGCTCCCAAGTACTAGAGTTACAGGTGCAGACCAACGTCCA	60
QY	61	GCTAATTTTGTATTTTATTTAGTAGAGACGGGGTTTCAACCGTGTTCAGAGCTGGTATCAA	120
Db	61	GCTAATTTTGTATTTTATTTAGTAGAGACGGGGTTTCAACCGTGTTCAGAGCTGGTATCAA	120
QY	121	CTTCCTGACCTCAAGTATCTGCTGCTCTAGAGCTCCCAAAATGCTAGATTAACAGTGG	180
Db	121	CTTCCTGACCTCAAGTATCTGCTGCTCTAGAGCTCCCAAAATGCTAGATTAACAGTGG	180
QY	181	AGTACACCGACACGAGCCCTTCTTTCAGTCTATACACTCTTTTGGCTATATTTGTATGAG	240
Db	181	AGTACACCGACACGAGCCCTTCTTTCAGTCTATACACTCTTTTGGCTATATTTGTATGAG	240
QY	241	AGCTTTATTTATTTAGGGGCACATPACTTTAAATGTTATGCTTATTTGATATGATCT	300
Db	241	AGCTTTATTTATTTAGGGGCACATPACTTTAAATGTTATGCTTATTTGATATGATCT	300
QY	301	GTCAATTATTAAGTCTGTATTCATTCGCGATATTCCTTTTCTAAATTTTCTCG	360
Db	301	GTCAATTATTAAGTCTGTATTCATTCGCGATATTCCTTTTCTAAATTTTCTCG	360
QY	361	AATGTGTCTGCTATTAACATPAGCACTGCGCTTTTAAATATGATTTTATAGGTATA	420
Db	361	AATGTGTCTGCTATTAACATPAGCACTGCGCTTTTAAATATGATTTTATAGGTATA	420
QY	421	TATTTTCTCTTTTATTTTATTTAAGTTTAAATGTTATGTTTCTTATTACTTAAAGTGG	480
Db	421	TATTTTCTCTTTTATTTTATTTAAGTTTAAATGTTATGTTTCTTATTACTTAAAGTGG	480

QY	481	GTGCTTAATAGGAGCAGCATATATCGGGCTGAGATATTTAAATCGAATCTCAAC	540
Db	481	GTGCTTAATAGGAGCAGCATATATCGGGCTGAGATATTTAAATCGAATCTCAAC	540
QY	541	CTTTTGTGTGGAAGGTTTAGGCCATTATACATTTAGTGTAAATTAATAGCATGGTTGATTT	600
Db	541	CTTTTGTGTGGAAGGTTTAGGCCATTATACATTTAGTGTAAATTAATAGCATGGTTGATTT	600
QY	601	GCTATACCACCTTTTTCATTTGTGTTTATATATGAGCCATCTTTTCATGTTCTTTTTCAT	660
Db	601	GCTATACCACCTTTTTCATTTGTGTTTATATATGAGCCATCTTTTCATGTTCTTTTTCAT	660
QY	661	CTTTGACCATTTTCTTTTAGTACGTAAATCTTTTGTGATTTTCAATTAATATCTATGGGCTT	720
Db	661	CTTTGACCATTTTCTTTTAGTACGTAAATCTTTTGTGATTTTCAATTAATATCTATGGGCTT	720
QY	721	TTTAGATTATACCTCTTAAAAATTTTTTTTTTTCGTTTATAGATTTATATATATACATCT	780
Db	721	TTTAGATTATACCTCTTAAAAATTTTTTTTTTTCGTTTATAGATTTATATATATACATCT	780
QY	781	TTAACTTATACACAGATTACCTTCAAAATATATTTTTCACAGCTCAAGTGAATGTAGAAC	840
Db	781	TTAACTTATACACAGATTACCTTCAAAATATATTTTTCACAGCTCAAGTGAATGTAGAAC	840
QY	841	CTTACAAAGATATATTTGCAATTTCTGTCGCTAATTTTATGACATGCTATATATACAT	900
Db	841	CTTACAAAGATATATTTGCAATTTCTGTCGCTAATTTTATGACATGCTATATATACAT	900
QY	901	AGGTTGTTGTTGTTGTTTATTTTACCTTATGCTGTGGCTGGGCTCAGCAAACTTTTCT	960
Db	901	AGGTTGTTGTTGTTGTTTATTTTACCTTATGCTGTGGCTGGGCTCAGCAAACTTTTCT	960
QY	961	GTAAGGGGCTAGATATGTAACAGGCATACCTTGGAGATATCTGTGGTTTGTTCCATACAC	1020
Db	961	GTAAGGGGCTAGATATGTAACAGGCATACCTTGGAGATATCTGTGGTTTGTTCCATACAC	1020
QY	1021	CACATTAATACAAATATATGACAGAAAGTGAATATCACTAAAGATGACACAAGTCTTT	1080
Db	1021	CACATTAATACAAATATATGACAGAAAGTGAATATCACTAAAGATGACACAAGTCTTT	1080
QY	1081	TGGCTTCCAGTGCATATATAAAGTTTGCTATATCTACACGTATGCTGTAAATGTGCA	1140
Db	1081	TGGCTTCCAGTGCATATATAAAGTTTGCTATATCTATACACGTATGCTGTAAATGTGCA	1140
QY	1141	ATAGTGTATATGCTAAAAAAACACATACCTTAATTTTAAATGCTTTATATCTAAAAAT	1200
Db	1141	ATAGTGTATATGCTAAAAAAACACATACCTTAATTTTAAATGCTTTATATCTAAAAAT	1200
QY	1201	GCTAACATCATTTTATGAGCATTCAGTGAATGTATCTTTTTCGTGTGGAAGGCTTTTC	1260
Db	1201	GCTAACATCATTTTATGAGCATTCAGTGAATGTATCTTTTTCGTGTGGAAGGCTTTTC	1260
QY	1261	TTATTTGATGACATGATCGGGGGGTACGGGCTGAAGCTTAAAGGCTGTGGAGTTTCTTA	1320
Db	1261	TTATTTGATGACATGATCGGGGGGTACGGGCTGTAAAGCTTAAAGGCTGTGGAGTTTCTTA	1320
QY	1321	AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCATTACATAAAGATTTCTCTTA	1380
Db	1321	AAACAACAGTGAAGATTGCAATATCAGTTGACTCTTCCATTACATAAAGATTTCTCTTA	1380
QY	1381	GTGTGTGATGCTTTTGTATAGCATTTTATGCAACGTAGAACCTTCTTGAAATTTGATCA	1440
Db	1381	GTGTGTGATGCTTTTGTATAGCATTTTATGCAACGTAGAACCTTCTTGAAATTTGATCA	1440
QY	1441	ATTCCTCTCAAAACCGGCTCTGTCTTAAACAACCTAAGTTAAATATATTCGAAATCAAT	1500
Db	1441	ATTCCTCTCAAAACCGGCTCTGTCTTAAACAACCTAAGTTAAATATATTCGAAATCAAT	1500
QY	1501	GTGTGCATTTCAACAATTTTCAAGTGTCTTCAACAGGAGTAGATTTCAATCTCATTTTCT	1560
Db	1501	GTGTGCATTTCAACAATTTTCAAGTGTCTTCAACAGGAGTAGATTTCAATCTCATTTTCT	1560
QY	1561	GAGATGGAATCTTTGCTCATTCATATAGAAATTTCTCATCTGTTCAAGTTTATCATG	1620

```
Db      1561 GAGATGAACTTTGCTGATCCATAGAGAAATTCCTATGCTGTTTATCAG 1620
      |||
Qy      1621 AGATTGACCAATACATAGTATGTTCTGAGGCTCACTTATTTATTCAGTTCTTT 1680
      |||
Db      1621 AGATTGACCAATACATAGTATGTTCTGAGGCTCACTTATTTATTCAGTTCTTT 1680
      |||
Qy      1681 GCTGTTTACCAACATCTGAGTTCTCTCTCCATGAAAGTGTGAACCTCCAGTCA 1740
      |||
Db      1681 GCTGTTTACCAACATCTGAGTTCTCTCTCCATGAAAGTGTGAACCTCCAGTCA 1740
      |||
Qy      1741 TCCATGAGGTTGGAATCCAGTCTCTTCCAATTCTCTTAATATTTATTTGACCTCC 1800
      |||
Db      1741 TCCATGAGGTTGGAATCCAGTCTCTTCCAATTCTCTTAATATTTATTTGACCTCC 1800
      |||
Qy      1801 CATGATCATGAAATGTTCTTTAATGCGACCTGGAAATGTAATCTTCCAAAAGGTTTC 1860
      |||
Db      1801 CATGATCATGAAATGTTCTTTAATGCGACCTGGAAATGTAATCTTCCAAAAGGTTTC 1860
      |||
Qy      1861 AATTACTAGTCCAGATCCATCCATCCAGAGATCCATTTCAATGCAATATAGCTT 1920
      |||
Db      1861 AATTACTAGTCCAGATCCATCCATCCAGAGATCCATTTCAATGCAATATAGCTT 1920
      |||
Qy      1921 TATGAAATGTAATTTCTTCAATATAAGCTTGAAGTGAATTAATCTCTGATCCATT 1980
      |||
Db      1921 TATGAAATGTAATTTCTTCAATATAAGCTTGAAGTGAATTAATCTCTGATCCATT 1980
      |||
Qy      1981 TCTGCMAAATAGATGTTGTG 2000
      |||
Db      1981 TCTGCMAAATAGATGTTGTG 2000
      |||

RESULT 2
AF222855      80622 bp      DNA      linear      PRI 17-JUL-2000
LOCUS      Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.
DEFINITION
ACCESSION      AF222855
VERSION      AF222855.1 GI:9246845
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
AUTHORS      Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
Irvine,D.V., Saferly,R. and Choo,K.H.
TITLE      The 10q25 neocentromere and its inactive progenitor have identical
primary nucleotide sequence: further evidence for epigenetic
modification
JOURNAL      Genome Res. 10 (6), 832-838 (2000)
PUBMED      10854414
REFERENCE
AUTHORS      Barry,A.E.
TITLE      Direct Subdivision
JOURNAL      Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
Parkville,, Melbourne, Victoria 3052, Australia
REMARK
FEATURES
source
1.80622
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25.2"
/clone="HC"
/repeat_region
/rpt_family="Alu"
complement(13..176)
repeat_region
/rpt_family="Alu"
complement(979..3564)
repeat_region
/rpt_family="Tigger1"
2967..3094
repeat_region
/rpt_family="Alu"
```

```
repeat_region /rpt_family="Alu"
complement(53270).53562)
repeat_region /rpt_family="Alu"
54261..54808
repeat_region /rpt_family="LTR12"
59460..60054
repeat_region /rpt_family="LTR12"
60349..60633
repeat_region /rpt_family="Alu"
complement(60669).60801)
repeat_region /rpt_family="MIR"
complement(61684).61852)
repeat_region /rpt_family="MERS"
63466..63753
repeat_region /rpt_family="Alu"
64063..64228
repeat_region /rpt_family="MERS"
64446..64521
repeat_region /rpt_family="MIR"
65794..65883
repeat_region /rpt_family="MERS"
66772..66951
repeat_region /rpt_family="MER20"
complement(67895).68416)
repeat_region /rpt_family="MER41"
complement(73192).73470)
repeat_region /rpt_family="Alu"
74086..74367
repeat_region /rpt_family="Alu"
75916..76183
repeat_region /rpt_family="Alu"
complement(77771).78059)
repeat_region /rpt_family="Alu"
```

## ORIGIN

```
Query Match 100.0%; Score 2000; DB 8; Length 80622;
Best Local Similarity 100.0%; Pred. No. 1.3e-300;
Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GAATTCCTCGGCTCAGCCCTCCCAAGTACTGAGGTTACAGGTCAGCCACACAGTCCA 60
DB 1 GAATTCCTCGGCTCAGCCCTCCCAAGTACTGAGGTTACAGGTCAGCCACACAGTCCA 60
OY 61 GCTAAATTTTGTATTTTGTAGTAGAGCGGGGTTTCAACCGTGTTCGACGGCTGTATCAA 120
DB 61 GCTAAATTTTGTATTTTGTAGTAGAGCGGGGTTTCAACCGTGTTCGACGGCTGTATCAA 120
OY 121 CTCCTGACCTCAAGTGATTCGCTGCTCAGCCTCCCAAAATGCTAGATTACAGGTG 180
DB 121 CTCCTGACCTCAAGTGATTCGCTGCTCAGCCTCCCAAAATGCTAGATTACAGGTG 180
OY 181 AATCACCAGCAGCAGCCCTTCTTTCAGTCTATTCACCTTTTGTATTTGTATTTGATG 240
DB 181 AATCACCAGCAGCAGCCCTTCTTTCAGTCTATTCACCTTTTGTATTTGTATTTGATG 240
OY 241 AGCTTATTTATTTAGGGGACATATCAATTTAAATTTGTATTTGTATTTAGATTGATCT 300
DB 241 AGCTTATTTATTTAGGGGACATATCAATTTAAATTTGTATTTGTATTTAGATTGATCT 300
OY 301 GTCATTATGAATGCTGTAATTCATTCCTGATAGTATTTCTTTCTTAATATTTTCTG 360
DB 301 GTCATTATGAATGCTGTAATTCATTCCTGATAGTATTTCTTTCTTAATATTTTCTG 360
OY 361 AATGTCGTGATTAATACATAGCACTGCGCTTTTAAATTTAGATTATTTATGTTATA 420
DB 361 AATGTCGTGATTAATACATAGCACTGCGCTTTTAAATTTAGATTATTTATGTTATA 420
OY 421 TATTTTCTTTTATTTTATTTTAAATTTAGTTAGTTATTTCTTATCTTAACTAAAGTG 480
DB 421 TATTTTCTTTTATTTTATTTTAAATTTAGTTAGTTATTTCTTATCTTAACTAAAGTG 480
OY 481 GTGCTTATAGCAGCATATATCTGGGCTTGTATTTATTTAATCTGATAATCTCAAC 540
DB 481 GTGCTTATAGCAGCATATATCTGGGCTTGTATTTATTTAATCTGATAATCTCAAC 540
```

```
DB 481 GTGCTTATAGCAGCATATATCTGGGCTTGTATTTATTTAATCTGATAATCTCAAC 540
OY 541 CTTTGTGTGGAGTGTGTAGGCCATTATCAATTTAGTGAATTAATAGACATGTTGATTT 600
DB 541 CTTTGTGTGGAGTGTGTAGGCCATTATCAATTTAGTGAATTAATAGACATGTTGATTT 600
OY 601 GCTATACCATCTTTTCAATTTGTTTATATAGGACCATCTTTTCATGTTCTTTTTCAT 660
DB 601 GCTATACCATCTTTTCAATTTGTTTATATAGGACCATCTTTTCATGTTCTTTTTCAT 660
OY 661 CTTTGACATTTTCTTTAGTACGTAATCTTTTGTATTTCAATTAATCATTTGCTT 720
DB 661 CTTTGACATTTTCTTTAGTACGTAATCTTTTGTATTTCAATTAATCATTTGCTT 720
OY 721 TTTAGTTATACCTTTAAATTTTCTTGTATTTATAGATTATTAATTAATATCATCT 780
DB 721 TTTAGTTATACCTTTAAATTTTCTTGTATTTATAGATTATTAATTAATATCATCT 780
OY 781 TTTAATTATCAAGATTACCTTCAAAATTTATTTTACAGCTCAAGCTAATGTAGAAC 840
DB 781 TTTAATTATCAAGATTACCTTCAAAATTTATTTTACAGCTCAAGCTAATGTAGAAC 840
OY 841 CTTACAGAGTATATTTTCAATTTCTGTCCTAAATTTTATGCTATGCTATTAATACAT 900
DB 841 CTTACAGAGTATATTTTCAATTTCTGTCCTAAATTTTATGCTATGCTATTAATACAT 900
OY 901 AGGTTGTGTGTTTGTATTTTATTTTACCTTAATTTGCTGTGGCTGAGCAAAATTTTCT 960
DB 901 AGGTTGTGTGTTTGTATTTTATTTTACCTTAATTTGCTGTGGCTGAGCAAAATTTTCT 960
OY 961 GTAAAGGCTGATATGATACAGCATACCTTGTGAATATCTGTGGTTGGTCCATACAC 1020
DB 961 GTAAAGGCTGATATGATACAGCATACCTTGTGAATATCTGTGGTTGGTCCATACAC 1020
OY 1021 CACAATTAATACAAATATGCAAGAGTGAATATCAATTAAGTGAATATCAATTAAGTGA 1080
DB 1021 CACAATTAATACAAATATGCAAGAGTGAATATCAATTAAGTGAATATCAATTAAGTGA 1080
OY 1081 TGGCTTCCAGTGCATATTAAGTTTGTCTTATATCACTAGTATCTGTTAAGTGTCA 1140
DB 1081 TGGCTTCCAGTGCATATTAAGTTTGTCTTATATCACTAGTATCTGTTAAGTGTCA 1140
OY 1141 ATAGTGTATGCTTAATAAAAAACATATCTTAATTTTAAATGCTTTATTTCTAATAAAT 1200
DB 1141 ATAGTGTATGCTTAATAAAAAACATATCTTAATTTTAAATGCTTTATTTCTAATAAAT 1200
OY 1201 GCTAACAATTTGAGGATTCAGAGTGTGAATCTTTTGTGCTGAGGAGGCTTTTC 1260
DB 1201 GCTAACAATTTGAGGATTCAGAGTGTGAATCTTTTGTGCTGAGGAGGCTTTTC 1260
OY 1261 TTAATGATGACTGATCGGGGGTCAAGTGTGAAGCTTAAGGCTGTGGCAGTTCTTA 1320
DB 1261 TTAATGATGACTGATCGGGGGTCAAGTGTGAAGCTTAAGGCTGTGGCAGTTCTTA 1320
OY 1321 AAAACAAGTGAAGTGAATATATCAATGATGCTCTTCTTCAATGAAGATTTCTCTTA 1380
DB 1321 AAAACAAGTGAAGTGAATATATCAATGATGCTCTTCTTCAATGAAGATTTCTCTTA 1380
OY 1381 GTGTCGATGCTTTTGTATGATATTTATGACAGTGAAGCTTTGAAAATTTGATCA 1440
DB 1381 GTGTCGATGCTTTTGTATGATATTTATGACAGTGAAGCTTTGAAAATTTGATCA 1440
OY 1441 ATCCCTCAAAACCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTTGTAATCA 1500
DB 1441 ATCCCTCAAAACCTGCTCTGCTTTAACAACCTAAGTTAATATAATATTTGTAATCA 1500
OY 1501 GTTGTCAATTTCAAAATTTTCAAGTGTCTTCAACAGAGTATGATTCATCTCATTTCT 1560
DB 1501 GTTGTCAATTTCAAAATTTTCAAGTGTCTTCAACAGAGTATGATTCATCTCATTTCT 1560
OY 1561 GAGATGGAATCTTTGCTATCAATAAGAAATTTCTCATCTGTTCAAGTTTATCAG 1620
DB 1561 GAGATGGAATCTTTGCTATCAATAAGAAATTTCTCATCTGTTCAAGTTTATCAG 1620
```

QY 1621 AGATTCAGCAATACAGTACGTCTTACAGGCTCAGCTTCACTTTAATTCAGTCTCTT 1680  
 |||||  
 Db 1621 AGATTCAGCAATACAGTACGTCTTACAGGCTCAGCTTCACTTTAATTCAGTCTCTT 1680  
 QY 1681 GCTGTTTCTACCAATCTGTGTTCTTCTTCCATTAAGTCTTGAACCTTCCAACTGA 1740  
 |||||  
 Db 1681 GCTGTTTCTACCAATCTGTGTTCTTCTTCCATTAAGTCTTGAACCTTCCAACTGA 1740  
 QY 1741 TCCATAGAGGTTGGATCCACTCTTCCAAATTCCTTTAATATTATTTTGAACCTCC 1800  
 |||||  
 Db 1741 TCCATAGAGGTTGGATCCACTCTTCCAAATTCCTTTAATATTATTTTGAACCTCC 1800  
 QY 1801 CATGAATCATGAATGTTCTTAATAGCACCCTGAAGTGAATCTTCCAAAGGTTTC 1860  
 |||||  
 Db 1801 CATGAATCATGAATGTTCTTAATAGCACCCTGAAGTGAATCTTCCAAAGGTTTC 1860  
 QY 1861 AATTACTAGTCCAGATCCATCCAGAGATCCACTTTCATGCGAGTTATAGCCT 1920  
 |||||  
 Db 1861 AATTACTAGTCCAGATCCATCCAGAGATCCACTTTCATGCGAGTTATAGCCT 1920  
 QY 1921 TATGAAATGATTTCTTCAATATAAGCTTGAAGTTGAAATTAATCTCTTGAATTCATTT 1980  
 |||||  
 Db 1921 TATGAAATGATTTCTTCAATATAAGCTTGAAGTTGAAATTAATCTCTTGAATTCATTT 1980  
 QY 1981 TCTGCAAAATAGATGTTGTG 2000  
 |||||  
 Db 1981 TCTGCAAAATAGATGTTGTG 2000

RESULT 3  
 AL355340 112442 bp DNA linear PRI 18-MAY-2005  
 LOCUS AL355340  
 DEFINITION Human DNA sequence from clone RP11-383C6 on chromosome 10. Contains the TRUB1 gene for Trub pseudouridine (psi) synthase homolog 1 (E. coli), a novel gene and a CpG island, complete sequence.  
 ACCESSION AL355340  
 VERSION AL355340.17 GI:21436504  
 KEYWORDS HTG; TRUB1.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 112442)  
 AUTHORS Leonamornleert,D.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
 COMMENT Clone requests: clonerequests@sanger.ac.uk  
 On Jun 17, 2002 this sequence version replaced gi:16416169.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-383C6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: vegas@sanger.ac.uk  
 -----  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >=

30), an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

source  
 1..112442  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="RZPD:RPCI18753C06383"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-383C6"  
 /clone\_1ib="RPCI-11.2"  
 2000  
 /note="Clone right end: RP11-106M7"  
 join(24099..24445,28551..28649,37000..37055,45632..45713,56274..56346,58041..58180,60235..60291,61029..63577)  
 /gene="TRUB1"  
 /locus\_tag="RP11-383C6.1-001"  
 join(24099..24445,28551..28649,37000..37055,45632..45713,56274..56346,58041..58180,60235..60291,61029..63577)  
 /gene="TRUB1"  
 /locus\_tag="RP11-383C6.1-001"  
 /product="Trub pseudouridine (psi) synthase homolog 1 (E. coli)"  
 /note="match: ESTs: A1208778.1 B0775335.1  
 match: CDNA: AF448144.1 BC030601.1"  
 join(24192..24445,28551..28649,37000..37055,45632..45713,56274..56346,58041..58180,60235..60291,61029..61285)  
 /gene="TRUB1"  
 /locus\_tag="RP11-383C6.1-001"  
 /standard\_name="OTTHUMP0000020554"  
 /note="match: protein: Q8MWH5 Q9D311"  
 /code\_start=1  
 /product="Trub pseudouridine (psi) synthase homolog 1 (E. coli)"  
 /protein\_id="CAH73219.1"  
 /db\_xref="GI:55663184"  
 /db\_xref="Genbank:16060"  
 /db\_xref="GOA:Q8MWH5"  
 /db\_xref="InterPro:IPR002501"  
 /db\_xref="InterPro:IPR004510"  
 /translation="MAASEAAVYSSPKTDTSPLETRAGTYAAMAPPSAARAAAYVAAARTGSEARVSKALATKLSLSGCVAVHPEKPTSAELNRLKEKILAEKMPSPFWTKRKQYTKLKGIGTLDASARGLVVGISGTMLTSMISGSRVYATIGELKATDITDSTYATTEKPYDKITQEDIEGILQRTGIMQVPPPLYSALKDGGQRLSTIMRGEVVEAKPARPVYVYSISLQKFPFTLDVCEGCGFYIRSLVDICKELSSCANVLELTRTKQGPFLTEHALPEDKMTIDIDINQSLHCSLRFPAELALKSKRPSNSQVLSCEYITLNEKRDDVYKTC"  
 join(24292..24445,28551..28649,37000..37055,45632..45906)  
 /gene="TRUB1"  
 /locus\_tag="RP11-383C6.1-002"  
 join(24292..24445,28551..28649,37000..37055,45632..45906)  
 /locus\_tag="RP11-383C6.1-001"  
 /locus\_tag="RP11-383C6.1-002"  
 /product="Trub pseudouridine (psi) synthase homolog 1 (E. coli)"  
 /note="match: ESTs: A1418848.1 A1422325.1 A1640751.1  
 BE858241.1"  
 61455..61460  
 /gene="TRUB1"  
 /locus\_tag="RP11-383C6.1-001"  
 61479  
 /gene="TRUB1"  
 /locus\_tag="RP11-383C6.1-001"  
 63558..63563  
 /gene="TRUB1"  
 /locus\_tag="RP11-383C6.1-001"  
 63582  
 join(80563..80691,81349..81444,82381..82499)  
 /locus\_tag="RP11-383C6.2-001"  
 join(80563..80691,81349..81444,82381..82499)



/locus\_tag="Rp11-3836c.2-001"  
/product="putative novel transcripc"  
/note="match: ESTs: AA74571.1"  
110443  
misc\_feature  
/note="Clone\_left\_end: Rp11-338111"

ORIGIN

Query Match 98.3%; Score 1966; DB 8; Length 112442;  
Best Local Similarity 99.9%; Pred. No. 2.1e-295;  
Matches 1999; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 1 GAATTCCTGCTCAAGCTCCCAAGTAGCTAGGTTACAGGTGCCAGCCACCAAGTCCA 60  
DB 68556 GAATTCCTGCTCAAGCTCCCAAGTAGCTAGGTTACAGGTGCCAGCCACCAAGTCCA 68615  
QY 61 GCTATTTTGTATTTTATAGTAGAGAGGGGTTTCAACCGTTTGGCAGGCTGTATCAAA 120  
DB 68616 GCTATTTTGTATTTTATAGTAGAGAGGGGTTTCAACCGTTTGGCAGGCTGTATCAAA 68675  
QY 121 CTCCTGACCTCAAGTAGCTGCTGCTGAGCTCCCAAAATGCTAGATTACAGGTGTG 180  
DB 68676 CTCCTGACCTCAAGTAGCTGCTGCTGAGCTCCCAAAATGCTAGATTACAGGTGTG 68735  
QY 181 AGTCACCGCACCGAGCCCTTCTTCAAGTCTACCTCTTTTGTCTATTTGTATGAG 240  
DB 68736 AGTCACCGCACCGAGCCCTTCTTCAAGTCTACCTCTTTTGTCTATTTGTATGAG 68795  
QY 241 AGCTTATTTATAGGGGGCATACATTTAAATGTTATGCTTATGATAGATTGATCT 300  
DB 68796 AGCTTATTTATAGGGGGCATACATTTAAATGTTATGCTTATGATAGATTGATCT 68855  
QY 301 GTCATTATGAAATGCTGATTCATTCCTGATAGTATTCCTTTCTTAATAATTTTCTG 360  
DB 68856 GTCATTATGAAATGCTGATTCATTCCTGATAGTATTCCTTTCTTAATAATTTTCTG 68914  
QY 361 AATGTCGTCTATTAAACATPAGCCACTGCTTTTAAATATGATTTTATGATATA 420  
DB 68915 AATGTCGTCTATTAAACATPAGCCACTGCTTTTAAATATGATTTTATGATATA 68974  
QY 421 TATTTTCTCTTTTCTTTTCTTTTAAATTAGATGTTATGTTTCTTAACTAAAGTGG 480  
DB 68975 TATTTTCTCTTTTCTTTTCTTTTAAATTAGATGTTATGTTTCTTAACTAAAGTGG 69034  
QY 481 GTGCTTATAGCAGCATATCTGGGCTTGTGATGATTTAACTGATATCTCAAC 540  
DB 69035 GTGCTTATAGCAGCATATCTGGGCTTGTGATGATTTAACTGATATCTCAAC 69094  
QY 541 CTTTTTGTGAGTGTTTAGGCCATTTAACATTTAGTGAATTAAGACATGTTGATTT 600  
DB 69095 CTTTTTGTGAGTGTTTAGGCCATTTAACATTTAGTGAATTAAGACATGTTGATTT 69154  
QY 601 GCTATPACATCTTTTCAATTTGTTTATGATGAGCATCTTTCAATGTTCTTTTCAAT 660  
DB 69155 GCTATPACATCTTTTCAATTTGTTTATGATGAGCATCTTTCAATGTTCTTTTCAAT 69214  
QY 661 CTTTGCACATTTCTTATAGTATGATCTTTTGTATTTGATTAATATCTATGGCTT 720  
DB 69215 CTTTGCACATTTCTTATAGTATGATCTTTTGTATTTGATTAATATCTATGGCTT 69274  
QY 721 TTTAGTTATACCTCTTAAATTTTCTTCTGTTTATGATGAGATTATATATACATCT 780  
DB 69275 TTTAGTTATACCTCTTAAATTTTCTTCTGTTTATGATGAGATTATATATACATCT 69334  
QY 781 TTTATCTTATACAGATTACTTCAATATGATTTTACAGCTCAAGTGAATGTAGAAC 840  
DB 69335 TTTATCTTATACAGATTACTTCAATATGATTTTACAGCTCAAGTGAATGTAGAAC 69394  
QY 841 CTTACAAGATATATTTTCAATTTCTGCTCCTAATTTTATGCTA-TGCTATATATACAT 899  
DB 69395 CTTACAAGATATATTTTCAATTTCTGCTCCTAATTTTATGCTAATGCTATATATACAT 69454  
QY 900 TAGGTTGTGTTGTTGTTTATACCTTATGCTGTGGCTGGGGTCAGCAAACTTTTCT 959  
DB TAGGTTGTGTTGTTGTTTATACCTTATGCTGTGGCTGGGGTCAGCAAACTTTTCT 959

DB 69455 TAGGTTGTGTTGTTGTTTATACCTTATGCTGTGGCTGGGGTCAGCAAACTTTTCT 69514  
QY 960 TGTAAAGGCTAGATATGATACAGCATACCTTGAAGATATCTGTGGTTTGTTCATACCA 1019  
DB 69515 TGTAAAGGCTAGATATGATACAGCATACCTTGAAGATATCTGTGGTTTGTTCATACCA 69574  
QY 1020 CCAATATATACAAATATGCAAGAAATGGAATATCAATTAAGATGTCACAAAGTCTT 1079  
DB 69575 CCAATATATACAAATATGCAAGAAATGGAATATCAATTAAGATGTCACAAAGTCTT 69634  
QY 1080 TTGGCTTCCAGTGATATTAAGTTTGTCTTATCTACACGTGATGCTGTTAAGTGTGC 1139  
DB 69635 TTGGCTTCCAGTGATATTAAGTTTGTCTTATCTACACGTGATGCTGTTAAGTGTGC 69694  
QY 1140 AATAGTGTATGCTTAAAAAAACATATACCTTAATTTTAAATGCTTTATTAATAAAAA 1199  
DB 69695 AATAGTGTATGCTTAAAAAAACATATACCTTAATTTTAAATGCTTTATTAATAAAAA 69754  
QY 1200 TGTCAATCATCATTTGACATTCAGTGAATTTGATCTTTTGTCTGTGGAAGTCTTTT 1259  
DB 69755 TGTCAATCATCATTTGACATTCAGTGAATTTGATCTTTTGTCTGTGGAAGTCTTTT 69814  
QY 1260 CTTATGATGACTGATGCGGGGTCAGGTCGAAGCTTAGGGTGGCTGGGCACTTCTT 1319  
DB 69815 CTTATGATGACTGATGCGGGGTCAGGTCGAAGCTTAGGGTGGCTGGGCACTTCTT 69874  
QY 1320 AAAACAAGTAGAATTTGCAATATCAGTGAATCTTCTCTTCAATGAAGAATTTCTCT 1379  
DB 69875 AAAACAAGTAGAATTTGCAATATCAGTGAATCTTCTCTTCAATGAAGAATTTCTCT 69934  
QY 1380 AGTGTGATGCTTTTGTATATGACATTTTATGCAAGTGAACCTTTGAAAAATTGGA-T 1438  
DB 69935 AGTGTGATGCTTTTGTATATGACATTTTATGCAAGTGAACCTTTGAAAAATTGGA-T 69994  
QY 1439 CAATCCTCTCAAAACCTGCTCTGCTTTAAACAACCTTAAGTAAATATATATTTGTAATCA 1498  
DB 69995 CAATCCTCTCAAAACCTGCTCTGCTTTAAACAACCTTAAGTAAATATATATTTGTAATCA 70054  
QY 1499 TTGTTGTCATTTCAACAATTTTCAAGTGTCTTCAACAGAGTAGATTCATCTATTTTC 1558  
DB 70055 TTGTTGTCATTTCAACAATTTTCAAGTGTCTTCAACAGAGTAGATTCATCTATTTTC 70114  
QY 1559 CTGAGATGAATCTTTGCTCATCCATPAGAGAAATTCCTCATCTGTTCAAGTTTATCA 1618  
DB 70115 CTGAGATGAATCTTTGCTCATCCATPAGAGAAATTCCTCATCTGTTCAAGTTTATCA 70174  
QY 1619 TGAATATGAGAAATPAGAGTATGCTTCAAGGCTCACTTCACTTTAATTCAGATTCCTC 1678  
DB 70175 TGAATATGAGAAATPAGAGTATGCTTCAAGGCTCACTTCACTTTAATTCAGATTCCTC 70234  
QY 1679 TTGCTGTTTCTACCAATCTGTGTTCTTCTCTCAATTGAAGTCTTGAACCTCTCAAGT 1738  
DB 70235 TTGCTGTTTCTACCAATCTGTGTTCTTCTCTCAATTGAAGTCTTGAACCTCTCAAGT 70294  
QY 1739 CATTCATGAGGGTGGAAATGCACTTCTTCAAAATCTCTGTAAATTTATATTTGACCT 1798  
DB 70295 CATTCATGAGGGTGGAAATGCACTTCTTCAAAATCTCTGTAAATTTATATTTGACCT 70354  
QY 1799 CCCATGAATCATGAATGTTCTTAATGGAACCTGGAATGTTGAATCTTCCAAAAGTTT 1858  
DB 70355 CCCATGAATCATGAATGTTCTTAATGGAACCTGGAATGTTGAATCTTCCAAAAGTTT 70414  
QY 1859 TCAATTTACTTATGTCAGATTCATCATCCATCAAGAGATCCATTTCAATGCAAGTTATAGC 1918  
DB 70415 TCAATTTACTTATGTCAGATTCATCATCCATCAAGAGATCCATTTCAATGCAAGTTATAGC 70474  
QY 1919 CTTATGGAATGATTTCTTCAATTAATAGCTTGAAGTTGAATTTACTCTTGATCAAT 1978  
DB 70475 CTTATGGAATGATTTCTTCAATTAATAGCTTGAAGTTGAATTTACTCTTGATCAAT 70534  
QY 1979 TTTCTGCAAAATAGATGTTGTG 2000  
DB 70535 TTTCTGCAAAATAGATGTTGTG 70556



Db	44361	GGTAAACACATCTTTTCATTGTTTTPANPANGGCACTTTTCATTGTTCTTTTTCAT	44302
Oy	661	CTTGAACCAATTTTCTTTAGTACGAACTACTTTTTTGTATTCTTAATATCTATGGCTT	720
Db	44301	CTTTGACCAATTTTCTTTAGTACGAACTACTTTTTTGTATTCTTAATATCTATGGCTT	44242
Oy	721	TTTAGTTAATCCGCTTAAATATTTTTTCTGTTTTATGAGATTTATATATATACATCT	780
Db	44241	TTTAGTTAATCCCTTAAATATTTTTTCTGTTTTATGAGATTTATATATATACATCT	44182
Oy	781	TTAATCTTATCAGAGATTACCTTCAAAATAGTATTTTATACAGCTCAAGGTATGTAGAAC	840
Db	44181	TTAATCTTATCAGAGATTACCTTCAAAATAGTATTTTATACAGCTCAAGGTATGTAGAAC	44122
Oy	841	CTTACAGAGTATATTTTCAATTTCTGTCTCTCTAATTTTTATGCTA-TGTCTATATACAT	899
Db	44121	CTTACAGAGTATATTTTCAATTTCTGTCTCTCTAATTTTTATGCTATGTCTATATACAT	44062
Oy	900	TAGT	959
Db	44061	TAGT	44002
Oy	960	TGTAAAGGGCTAGATAGTACAGGCAATCTTGAGATACGTGTGGTGTGTGTGTGTGTGT	1019
Db	44001	TGTAAAGGGCTAGATAGTACAGGCAATCTTGAGATACGTGTGGTGTGTGTGTGTGTGT	43942
Oy	1020	CCACAATATATCAAAATATGCAAGAAGTGATATCACAAATAAAGTAGTCACACAAGTCTT	1079
Db	43941	CCACAATATATCAAAATATGCAAGAAGTGATATCACAAATAAAGTAGTCACACAAGTCTT	43882
Oy	1080	TTGGCTTCCAGTGCATATAAAGTTTGTCTTATACACCTGATAGTCTGTAAAGTGTG	1139
Db	43881	TTGGCTTCCAGTGCATATAAAGTTTGTCTTATACACCTGATAGTCTGTAAAGTGTG	43822
Oy	1140	AATAGTGTATGTCTTAAAAAAACACATACCTTAATTTTAAAAATCTTTATCTAAAAAA	1199
Db	43821	AATAGTGTATGTCTTAAAAAAACACATACCTTAATTTTAAAAATCTTTATCTAAAAAA	43762
Oy	1200	TGTAAACATCATTTGAGCATTCAGTAGTGTATCTTTTGTGTGTGTGTGTGTGTGT	1259
Db	43761	TGTAAACATCATTTGAGCATTCAGTAGTGTATCTTTTGTGTGTGTGTGTGTGTGT	43702
Oy	1260	CTTATTTGATGACTGATTCGGGGGTCAAGTGTCTGAAGCTTAAGGTGTGTGTGTGTGT	1319
Db	43701	CTTATTTGATGACTGATTCGGGGGTCAAGTGTCTGAAGCTTAAGGTGTGTGTGTGTGT	43642
Oy	1320	AAAAACAAGTGAAGATTGCAATATCAGTTGACCTTCCCTTCATGAAGAAATTCCTCT	1379
Db	43641	AAAAACAAGTGAAGATTGCAATATCAGTTGACCTTCCCTTCATGAAGAAATTCCTCT	43582
Oy	1380	AGTGTGTGATGCTTTTGTATAGCATTTTATATGACAGTAGAACTTCTTTGAAATTTGA-T	1438
Db	43581	AGTGTGTGATGCTTTTGTATAGCATTTTATATGACAGTAGAACTTCTTTGAAATTTGA	43522
Oy	1439	CAATCTCTCAAAACCTGTCTGTCTTTTAAACAACCTAAGTTAATATATATTTCTGAATCA	1498
Db	43521	CAATCTCTCAAAACCTGTCTGTCTTTTAAACAACCTAAGTTAATATATATTTCTGAATCA	43462
Oy	1499	TTGTGTGATTTTGAACAATTTTCAAGTCTTCAACGAGAGTAGATCCATCTCATTTTC	1558
Db	43461	TTGTGTGATTTTGAACAATTTTCAAGTCTTCAACGAGAGTAGATCCATCTCATTTTC	43402
Oy	1559	CTGAGATGGAATCTTTGTCTCATCATTAAGAAGAAATTCCTCATCTGTCAAGTTTATCA	1618
Db	43401	CTGAGATGGAATCTTTGTCTCATCATTAAGAAGAAATTCCTCATCTGTCAAGTTTATCA	43342
Oy	1619	TGAGATTTGACGAATATACATGATGCTTCAAGGCTCACTTCACTTTTAAATTCAGATTC	1678
Db	43341	TGAGATTTGACGAATATATACATGATGCTTCAAGGCTCACTTCACTTTTAAATTCAGATTC	43282
Oy	1679	TTGTCTTTTCTACACATCTGTGTGTCTTCTCTCATTTGAAGTCTTGAACCTCTCCAAAT	1738
Db	43281	TTGTCTTTTCTACACATCTGTGTGTCTTCTCTCATTTGAAGTCTTGAACCTCTCCAAAT	43222

Qy	1739	CATCCATGAGGGTTGGAAATCGATCTTCCTCCAAATTCGTGTAAATTTATATTTTGACCT	1798
Db	43221	CATCCATGAGGGTTGGAAATCGATCTTCCTCCAAATTCGTGTAAATTTATATTTTGACCT	43162
Qy	1799	CCCATGATCATGAATGTTCTTAATGAGCACCTGGAAATGATGAAATCCTTTCCAAAGTGT	1858
Db	43161	CCCATGATCATGAATGTTCTTAATGAGCACCTGGAAATGATGAAATCCTTTCCAAAGTGT	431020
Qy	1859	TCATTTACTTAGTGCAGATCCATCCATCCAGAGATCCACTTTCGAATGCCAGTTATAGC	1918
Db	43101	TCATTTACTTAGTGCAGATCCATCCATCCAGAGATCCACTTTCGAATGCCAGTTATAGC	430424
Qy	1919	CTTAATGGAATGTATTTCTTCAATTAATAGGCTTGAAGTTGAATTAATCTCTTGATCCAT	1978
Db	43041	CTTAATGGAATGTATTTCTTCAATTAATAGGCTTGAAGTTGAATTAATCTCTTGATCCAT	429822
Qy	1979	TTTCTGCAAAATGATGTTGTG 2000	
Db	42981	TTTCTGCAAAATGATGTTGTG 42960	

## RESULT 5

LOCUS	AX033912	41008 bp	DNA	linear	PAT 21-SEP-2000
DEFINITION	Sequence 4 from Patent W09851790.				
ACCESSION	AX033912				
VERSION	AX033912.1	GI:10280480			

**SOURCE**

unclassified sequences.

## REFERENCE

**AUTHORS** Cancilla, M.R., Choo, K.H. and Du, S.D.  
**TITLE** A novel nucleic acid molecule  
**JOURNAL** Patent: WO 9851790-A 4 19-NOV-1998;

```
source
1. .41008
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
```

## ORIGIN

Query Match      80.6%;    Score 1611.4;    DB 6;    Length 41008;

Best Local Similarity 97.4%; Pred. No. 2e-240;  
Matches 166; Conservative 2; Mismatches 40; Indels 3; Gaps 3;

292 GATTCATCTGTCATTATGATGTCGTATTCATTCCTGATAGTATTTCTTTCTAAAT 351

Db 2 GGTGATTTGNNATAAGGAGTTTGAATCAATCCGGAGCAATTTTTTAAAAA 61

352 ATTTTCTGAATGTCTGCTATTACATAGCCACTCGCTTTTAAAAATTGATATTT 411

Db 62 ATTTTTCGACG-GTTTCGTAAATAAAARCCAAATTGGTTTAAATAAGGATTT 120

412 TATGGTATATATTTTCCTTTTTTTTTTTTAAAGTTTAGAGCTATGTTCCCTAAC 471

Db 121 TATGGAAAAATTTCCCTTTTTTTTTTAAGTTTAGAGCTATGTTCCCTATAC 180

472 TTAAGTGGGTCTTATAGCAGCATATATCTGGCTTGATGATTATTAACTGAT 531

Db 181 TTAAGTGGGTCTTATAGCAGCATATATCGGCTTGATGATTTAATCTGAT 240

532 AATCTCAACCTTTTGTGGAGTGTTAGGCCATTTCATTAGTGAATTATAGACATG 591

Db 241 AATCTAACCTTTTGTGGAGTGTAGGCATTACATTACTAGTAATTATAGACATG 300

592 GTTGAATTGCTATACCATCTTTTCATTGTTTATATGAGGCATCTTTTCATTGTC 651

Db 301 GTTGAATTTGGCTATACCATCTTTCAATTTGTTATATGTGAGCATCTTTCAATGTTTC 360

QY	552	TTTTTCAATCTTGACCAATTTCTCTTATGACGAATACCTTTTGTGATTTCAATATATC	711
Db	361	TTTTTTCATCTTGACCAATTTCTTTAGTACGAATACCTTTTGTATTTCAATATATC	420
QY	712	TATTGGCTTTTAGTTATACCTCTTAAATTTTTTTTCTGTTTATGTAGATTTTATA	771
Db	421	TATTGGCTTTTAGTTATACCTCTTAAATTTTTTTTCTGTTTATGTAGATTTTATA	480
QY	772	TATATCTTTTATCTTATACAGATATACCTTCAATAGTATTTTACGACTCAAGTGTA	831
Db	481	TATATCTTTTATCTTATACAGATATACCTTCAATAGTATTTTACGACTCAAGTGTA	540
QY	832	TGTGAAACCTTACAGAGTATATTTTCAATTCTGCTCCCAATTTTATATGTCTA	890
Db	541	TGTGAAACCTTACAGAGTATATTTTCAATTCTGCTCCCAATTTTATCTATGTCT	600
QY	891	ATAATATACATTAAGTTTGTGTGTGTGTTTATACCTTATCTGTGTGGCTGGACGA	950
Db	601	ATAATATACATTAAGTTTGTGTGTGTGTTTATACCTTATCTGTGTGGCTGGACGA	660
QY	951	AACATTTTCTGTAAAGGGCTAGATATGACAGGCATACCTTGAGATCTGTGGGTTGGT	1010
Db	661	AACATTTTCTGTAAAGGGCTAGATATGACAGGCATACCTTGAGATCTGTGGGTTGGT	720
QY	1011	TCCATATACACACATATATCAAAATATGCAAGAAATGATATACAAATAAAGTAGTAC	1070
Db	721	TCCATATACACACATATATCAAAATATGCAAGAAATGATATACAAATAAAGTAGTAC	780
QY	1071	ACAAGTCTTTTGGCTTCCAGTGCATATAAAGTTTCTTATATACACTGTATGTCT	1130
Db	781	ACAAGTCTTTTGGCTTCCAGTGCATATAAAGTTTCTTATATACACTGTATGTCT	840
QY	1131	TAACTGTGCATATAGTTATATGCTTAAAAAAACATACCTTAAATTTTAAATGCTTAT	1190
Db	841	TCAGTGTGCATATAGTTATATGCTTAAAAAAACATATCTTAAATTTTAAATGCTTAT	900
QY	1191	ACTTAAAAATGCTTAAACATCATTTTGAGCATTCAGTGAAGTTGTATCTTTTGTGTGTGA	1250
Db	901	ACTTAAAAATGCTTAAACATCATTTTGAGCATTCAGTGAAGTTGTATCTTTTGTGTGTGA	960
QY	1251	AGGACTTTTCTTATGATGATCGATTCGGGGGTGAGGTGCTGATGATTAAGGTGTGG	1310
Db	961	AGGACTTTTCTTATGATGATCGATTCGGGGGTGAGGTGCTGATGATTAAGGTGTGG	1020
QY	1311	CAGTTTCTTAAACAACAGTGAAGATGCAATATCAGTTGACTCTTCTTTCATGAAAGA	1370
Db	1021	CAGTTTCTTAAACAACAGTGAAGATGCAATATCAGTTGACTCTTCTTTCATGAAAGA	1080
QY	1371	TTTTCTCTCTAGTGTGTATGCTTTTGTATAGCATTTTATGCAACAGTGAACCTTCTTGA	1430
Db	1081	TTTTCTCTCTAGTGTGTATGCTTTTGTATAGCATTTTATGCAACAGTGAACCTTCTTGA	1140
QY	1431	AATTGGA-TCAATCCTCTCAAAACCTGCTGCTTTTAAACAACCTTAAGTATATATATTT	1489
Db	1141	AATTGGAATCAATCCTCTCAAAACCTGCTGCTTTTAAACAACCTTAAGTATATATATTT	1200
QY	1490	CTGATATCAATGTTGTCAATTTCAACAATTTTCAAGTGTCTTCAACAGAGATAGATTCCA	1549
Db	1201	CTGATATCAATGTTGTCAATTTCAACAATTTTCAAGTGTCTTCAACAGAGATAGATTCCA	1260
QY	1550	TTCTCATTTCTGAGATGGAATCTTTGTCTATCATCAATAAGAAAGAAATTCCTCATCTTTCAA	1609
Db	1261	TTCTCATTTCTGAGATGGAATCTTTGTCTATCATCAATAAGAAAGAAATTCCTCATCTTTCAA	1320
QY	1610	GTTTTATATCATAGATGAGAAATATACATATATGCTTCAAGGCTCTCATCTTTTATTT	1669
Db	1321	GTTTTATATCATAGATGAGAAATATACATATATGCTTCAAGGCTCTCATCTTTTATTT	1380
QY	1670	CCAGTTTCTTGTCTGTTTCTTACCAACATGTGTGTTCTTCTCTCATTTGAAGCTTTGAAC	1729
Db	1381	CCAGTTTCTTGTCTGTTTCTTACCAACATGTGTGTTCTTCTCTCATTTGAAGCTTTGAAC	1440
QY	1730	TTCTCAAGTATCCATAGAGGTTGGAATCGACTTCTTCAAAATTCCTGTTATATATTTATA	1789

Db	1441	TTCTCCAGTCATTCATBAGGGCTGGATTCGACTTTCTTCGAAATTCCTGTTAAATTTTATA	1500		
Qy	1790	TTTTCACCTCCCATGGAATCATGATGTTCTTAAATGGCACTGGAAATGGTAATCCTTCC	1849		
Db	1501	TTTTCGACCTCCCATGGAATCATGATGTTCTTAAATGGCACTGGAAATGGTAATCCTTCC	1560		
Qy	1850	AAAAGGTTTTCATTTACTTAGTCCAGATCCATTCATCCAGAGATTCACCTTTCATGCC	1909		
Db	1561	AAAAGGTTTTCATTTACTTAGTCCAGATCCATTCATCCAGAGATTCACCTTTCATGCC	1620		
Qy	1910	AGTATATGCGCTTATGGAATGTATTTCTTCAATATAAGCTTGAAAGTTGAAATTAATCC	1969		
Db	1621	AGTATATGCGCTTATGGAATGTATTTCTTCAATATAAGCTTGAAAGTTGAAATTAATCC	1680		
Qy	1970	TTGATTCATTTTCTTCGAAATAGATGTTTGG	2000		
Db	1681	TTGATTCATTTTCTTCGAAATAGATGTTTGG	1711		
RESULT 6					
AF222856		69058 bp	DNA	linear	PRI 17-JUL-2000
LOCUS					
DEFINITION	Homo sapiens clone Pnc chromosome 10 map 10q25.2 genomic sequence.				
ACCESSION	AF222856				
VERSION	AF222856.1	GI:9246846			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;				
	Hominidae; Homo.				
REFERENCE	1 (bases 1 to 69058)				
AUTHORS	Barry,A.B., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M., Irvine,D.V., Saffey,R. and Choo,K.H.				
TITLE	The 10q25 nucleotide sequence and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification				
JOURNAL	Genome Res. 10 (6), 832-838 (2000)				
PUBMED	10854414				
REFERENCE	2 (bases 1 to 69058)				
AUTHORS	Barry,A.B.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville, Melbourne, Victoria 3052, Australia				
REMARK	Human genomic sequence from 10q25.2, clone1b=PNC				
FEATURES	Location/Qualifiers				
SOURCE	1. 69058				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	/chromosome="10"				
	/map="10q25.2"				
	/clone="Pnc"				
	complement(598..3183)				
repeat_region	/rpt_family="Tigster1"				
repeat_region	2587..2715				
repeat_region	/rpt_family="Alu"				
repeat_region	3283..3553				
repeat_region	/rpt_family="Alu"				
repeat_region	complement(4548..4809)				
repeat_region	/rpt_family="Alu"				
repeat_region	complement(5919..6211)				
repeat_region	/rpt_family="Alu"				
repeat_region	6257..6599				
repeat_region	/rpt_family="MLT2B2"				
repeat_region	complement(8848..9123)				
repeat_region	/rpt_family="Alu"				
repeat_region	complement(10913..11198)				
repeat_region	/rpt_family="Alu"				
repeat_region	complement(12062..12313)				
repeat_region	/rpt_family="Alu"				

```
repeat_region complement(13957. .14236)
/rpt_family="Alu"
repeat_region 1713. .17424
/rpt_family="Alu"
repeat_region complement(19882. .20197)
/rpt_family="Alu"
repeat_region complement(22664. .22956)
/rpt_family="Alu"
repeat_region complement(23935. .24113)
/rpt_family="MIR"
repeat_region 24585. .24884
/rpt_family="Alu"
repeat_region complement(25618. .25973)
/rpt_family="MER1"
repeat_region 26735. .26970
/rpt_family="MER2"
repeat_region 27173. .27449
/rpt_family="Alu"
repeat_region 27526. .27796
/rpt_family="Alu"
repeat_region 28702. .28978
/rpt_family="Alu"
repeat_region complement(30658. .30736)
/rpt_family="MLT1"
repeat_region complement(31460. .31993)
/rpt_family="MER1"
repeat_region complement(32130. .32248)
/rpt_family="MLT1"
repeat_region 32685. .32961
/rpt_family="Alu"
repeat_region 33113. .33403
/rpt_family="Alu"
repeat_region complement(34169. .34233)
/rpt_family="MIR"
repeat_region 34332. .34639
/rpt_family="Alu"
repeat_region complement(35993. .36081)
/rpt_family="MIR"
repeat_region complement(38438. .38721)
/rpt_family="Alu"
repeat_region 39081. .39333
/rpt_family="Alu"
repeat_region complement(42817. .42877)
/rpt_family="MIR"
repeat_region complement(44845. .45119)
/rpt_family="Alu"
repeat_region 45532. .45665
/rpt_family="MIR"
repeat_region complement(46257. .46490)
/rpt_family="Alu"
repeat_region 48262. .48578
/rpt_family="Alu"
repeat_region complement(48687. .50337)
/rpt_family="L1"
repeat_region complement(50349. .50611)
/rpt_family="Alu"
repeat_region complement(50670. .50831)
/rpt_family="L1"
repeat_region complement(51085. .51374)
/rpt_family="Alu"
repeat_region complement(52905. .53197)
/rpt_family="Alu"
repeat_region 53896. .54467
/rpt_family="LTR12"
repeat_region 59129. .59723
/rpt_family="LTR12"
repeat_region 60018. .60302
/rpt_family="Alu"
repeat_region complement(60338. .60470)
/rpt_family="MIR"
repeat_region complement(61353. .61521)
/rpt_family="MER5"
repeat_region 63135. .63416
```

```
repeat_region /rpt_family="Alu"
repeat_region 63724. .63889
/rpt_family="MER5"
repeat_region 64107. .64182
/rpt_family="MIR"
repeat_region 65459. .65548
/rpt_family="MER5"
repeat_region 66439. .66618
/rpt_family="MER20"
repeat_region complement(67562. .68084)
/rpt_family="MER41"

ORIGIN
Query Match 79.6%; Score 1592.8; DB 8; Length 69058;
Best Local Similarity 99.8%; Pred. No. 1.2e-237;
Matches 1616; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 383 CCACTCTGGCTTTTAAATAGATTTTANAGATATATTTCCCTTTTCTTTT 442
DB 1 CCACTCTGGCTTTTAAATAGATTTTANAGATATATTTCCCTTTTCTTTT 60
QY 443 TAAATTTAGATGTTATGTTCTTATCTTAAAGTGGGTCTTATAGCAGCATATAT 502
DB 61 TAAATTTAGATGTTATGTTCTTATCTTAAAGTGGGTCTTATAGCAGCATATAT 120
QY 503 CTGGCTCTTGATGTTATTTATCTGATTAATCTCAACCTTTTGTGAGTGTAGGC 562
DB 121 CTGGCTCTTGATGTTATTTATCTGATTAATCTCAACCTTTTGTGAGTGTAGGC 180
QY 563 CATTTACATTTAGTGTATTTATAGACATGTTGATTTGCTATACCATTTTCACTTGT 622
DB 181 CATTTACATTTAGTGTATTTATAGACATGTTGATTTGCTATACCATTTTCACTTGT 240
QY 623 TTTATATGAGCCATCTTTTCATGTTCTTTTTCATCTTTGACCATTTTCTTATAGTAC 682
DB 241 TTTATATGAGCCATCTTTTCATGTTCTTTTTCATCTTTGACCATTTTCTTATAGTAC 300
QY 683 TGAATACCTTTTGTGATTTCAATTAATCTAATGCTTTTATAGTTAACTCTTAAATTT 742
DB 301 TGAATACCTTTTGTGATTTCAATTAATCTAATGCTTTTATAGTTAACTCTTAAATTT 360
QY 743 TTTTTCCTGTTTATGATGATTTATATATATATATATCTTTTAACTTATACGATATACCT 802
DB 361 TTTTTCCTGTTTATGATGATTTATATATATATATATATCTTTTAACTTATACGATATACCT 420
QY 803 CAATATGATTTTACAGCTCAAGTGAATGAGAACTTACAGAGATATTTTCAAT 862
DB 421 CAATATGATTTTACAGCTCAAGTGAATGAGAACTTACAGAGATATTTTCAAT 480
QY 863 TCTGTCTCCATATTTTATGCTA-TGTCTATATATCATTAAGTTTGTGTGTTT 921
DB 481 TCTGTCTCCATATTTTATGCTATGTCTATATATCATTAAGTTTGTGTGTTT 540
QY 922 TACCTTATTTGCTGTGGCTGGGTGACGAACATTTTCTGTAAGGCTGATATATACAG 981
DB 541 TACCTTATTTGCTGTGGCTGGGTGACGAACATTTTCTGTAAGGCTGATATATACAG 600
QY 982 GCATACCTTGAAGATACGTGGGTGTTGGTTCATACCAACCAATATATCAATATGCA 1041
DB 601 GCATACCTTGAAGATACGTGGGTGTTGGTTCATACCAACCAATATATGCA 660
QY 1042 GAAGTGATATCACATTAAGTGAATGATGACACAAAGTCTTTGGCTTCCAGTGCATATAA 1101
DB 661 GAAGTGATATCACATTAAGTGAATGATGACACAAAGTCTTTGGCTTCCAGTGCATATAA 720
QY 1102 AGTTTGTCTTATATCACTGATGATCTGTTAAGTGTGCAATAGTGTATGCTTAAAAA 1161
DB 721 AGTTTGTCTTATATCACTGATGATCTGTTAAGTGTGCAATAGTGTATGCTTAAAAA 780
QY 1162 CACATACCTTATTTTAAATGCTTATATACATTAATGCTATACATCATTTGAGCATTT 1221
DB 781 CACATACCTTATTTTAAATGCTTATATACATTAATGCTATACATCATTTGAGCATTT 840
```

QY 1222 CAGTAGGTTGATATCTTTTTCGTGAGGAGGCTTTTCTTATGATGACTGCGGG 1281  
| | | | |  
Db 841 CAGTAGGTTGATCTTTTTCGTGAGGAGGCTTTTCTTATGATGACTGCGGG 900  
| | | | |  
QY 1282 TCAGGTGCTGAAGCTTGAAGTGGCTGTGACAGTTCTTAAACACAGTGAAGATTGCA 1341  
| | | | |  
Db 901 TCAGGTGCTGAAGCTTGAAGTGGCTGTGACAGTTCTTAAACACAGTGAAGATTGCA 960  
| | | | |  
QY 1342 TATCAGGTGCTTCTTCTTCAATGAAGATTTCTCTGATGAGTGAAGCTTTTGATG 1401  
| | | | |  
Db 961 TATCAGGTGCTTCTTCTTCAATGAAGATTTCTCTGATGAGTGAAGCTTTTGATG 1020  
| | | | |  
QY 1402 CATTTTATGACAGTGAAGCTTCTTGAAGATTTGA-TCGATCTCTCAACCCCTGCT 1460  
| | | | |  
Db 1021 CATTTTATGACAGTGAAGCTTCTTGAAGATTTGAAGTGAATCTCTCAACCCCTGCT 1080  
| | | | |  
QY 1461 GCTTTTACACCTAAGTTAATAATATTTCTGAATCATTGTGTTGCTTCAACAAATTT 1520  
| | | | |  
Db 1081 GCTTTTACACCTAAGTTAATAATATTTCTGAATCATTGTGTTGCTTCAACAAATTT 1140  
| | | | |  
QY 1521 CACAGTGTCTTACACGAGAGTGAATCCATCTCATCTTCTGAGATGAATCTTTGCTCAT 1580  
| | | | |  
Db 1141 CACAGTGTCTTACACGAGAGTGAATCCATCTCATCTTCTGAGATGAATCTTTGCTCAT 1200  
| | | | |  
QY 1581 CCATAGAAGAAATTCCTCATCTGTTCAAGTTTATCATGAGATTGACCAATPACAGTCA 1640  
| | | | |  
Db 1201 CCATAGAAGAAATTCCTCATCTGTTCAAGTTTATCATGAGATTGACCAATPACAGTCA 1260  
| | | | |  
QY 1641 TGTCTTCAAGGCTCACTTCACTTTTAATTCAGTTCTCTTGTGTTCTTACACATCTGT 1700  
| | | | |  
Db 1261 TGTCTTCAAGGCTCACTTCACTTTTAATTCAGTTCTCTTGTGTTCTTACACATCTGT 1320  
| | | | |  
QY 1701 GGTCTCTTCTTCACTTGAAGTCTTGAACCTCTCCAGATCATCATGAGGCTTGAATGCA 1760  
| | | | |  
Db 1321 GGTCTCTTCTTCACTTGAAGTCTTGAACCTCTCCAGATCATCATGAGGCTTGAATGCA 1380  
| | | | |  
QY 1761 CTCTTCCAAATTCGTTAATATTTATTTTGAACCTCCAGATCATCATGAGGCTTGAATGCT 1820  
| | | | |  
Db 1381 CTCTTCCAAATTCGTTAATATTTATTTTGAACCTCCAGATCATCATGAGGCTTGAATGCT 1440  
| | | | |  
QY 1821 AATGACCTGAGATGATGATCTTCCAAAGGTTTCAATTTACTTATGTCAGATCTC 1880  
| | | | |  
Db 1441 AATGACCTGAGATGATGATCTTCCAAAGGTTTCAATTTACTTATGTCAGATCTC 1500  
| | | | |  
QY 1881 ATCCATCCAGAGATCCACTTCAATGCCAGTTATGCTTATGCAATGATTTCTTCA 1940  
| | | | |  
Db 1501 ATCCATCCAGAGATCCACTTCAATGCCAGTTATGCTTATGCAATGATTTCTTCA 1560  
| | | | |  
QY 1941 TATATAGGCTTGAAGTTGAATTTACTCTTGATCCATTTTCTGCAAAATPACATGTTG 2000  
| | | | |  
Db 1561 TATATAGGCTTGAAGTTGAATTTACTCTTGATCCATTTTCTGCAAAATPACATGTTG 1620  
| | | | |

RESULT 7  
AF222854 80202 bp DNA linear PRI 17-JUL-2000  
DEFINITION Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.  
ACCESSION AF222854  
VERSION AF222854.1 GI:9246844  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Hom sapiens  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE  
AUTHORS Barry A.E., Bateman, M., Howman, E.V., Cancellia, M.R., Tainon, K.M.,  
TITLE Irvine, D.V., Saferly, R. and Cho, K.H.  
The 10q25 neocentromere and its inactive progenitor have identical  
primary nucleotide sequence: further evidence for epigenetic  
modification  
JOURNAL Genome Res. 10 (6), 832-838 (2000)  
PUBMED 10854414

REFERENCE 2 (bases 1 to 80202)  
AUTHORS Barry A.E.  
TITLE Direct Submission  
JOURNAL Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch  
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,  
Parkville, Melbourne, Victoria 3052, Australia  
REMARK Human genomic sequence from 10q25.2, clone1b=NC, second release  
FEATURES  
source  
1. 80202  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10q25.2"  
/clone="NC"  
repeat\_region  
/rpt\_family="Tiger1"  
2587..2715  
/rpt\_family="Alu"  
repeat\_region  
3283..3553  
/rpt\_family="Alu"  
repeat\_region  
/rpt\_family="Alu"  
complement(4548..4809)  
/rpt\_family="Alu"  
repeat\_region  
complement(5919..6211)  
/rpt\_family="Alu"  
6257..6599  
/rpt\_family="MLT2B2"  
complement(8848..9123)  
/rpt\_family="Alu"  
repeat\_region  
complement(10913..11198)  
/rpt\_family="Alu"  
repeat\_region  
complement(12062..12313)  
/rpt\_family="Alu"  
repeat\_region  
complement(13957..14236)  
/rpt\_family="Alu"  
17133..17424  
/rpt\_family="Alu"  
repeat\_region  
complement(19882..20197)  
/rpt\_family="Alu"  
repeat\_region  
complement(22664..22956)  
/rpt\_family="Alu"  
repeat\_region  
complement(23935..24113)  
/rpt\_family="MIR"  
24585..24884  
/rpt\_family="Alu"  
repeat\_region  
complement(25618..25973)  
/rpt\_family="THE1"  
26732..26970  
/rpt\_family="MER2"  
27173..27449  
/rpt\_family="Alu"  
repeat\_region  
27526..27996  
/rpt\_family="Alu"  
repeat\_region  
28702..28978  
/rpt\_family="Alu"  
repeat\_region  
complement(30668..30736)  
/rpt\_family="MLT1"  
complement(31460..31993)  
/rpt\_family="MER1"  
complement(32130..32248)  
/rpt\_family="MLT1"  
32685..32961  
/rpt\_family="Alu"  
repeat\_region  
33113..33403  
/rpt\_family="Alu"  
repeat\_region  
complement(34169..34233)  
/rpt\_family="MIR"  
34332..34639  
/rpt\_family="Alu"  
repeat\_region  
complement(35993..36081)  
/rpt\_family="MIR"  
complement(38438..38721)  
/rpt\_family="Alu"  
repeat\_region

repeat_region	39081. . 39333	/rpt_family="Alu"	
repeat_region		complement(42817)	.42877)
repeat_region		/rpt_family="MIR"	
repeat_region		complement(44945)	.45119)
repeat_region	45537. .45665	/rpt_family="Alu"	
repeat_region		/rpt_family="MIR"	
repeat_region		complement(46257)	.46490)
repeat_region	48267. .48578	/rpt_family="Alu"	
repeat_region		complement(48687)	.50337)
repeat_region		/rpt_family="L1"	
repeat_region		complement(50349)	.50611)
repeat_region		/rpt_family="Alu"	
repeat_region		complement(50670)	.50831)
repeat_region		/rpt_family="L1"	
repeat_region		complement(51085)	.51374)
repeat_region		/rpt_family="Alu"	
repeat_region		complement(52905)	.53197)
repeat_region		/rpt_family="Alu"	
repeat_region	53895. .54467		
repeat_region		/rpt_family="LTR12"	
repeat_region	59129. .59719		
repeat_region		/rpt_family="LTR12"	
repeat_region	60018. .60302		
repeat_region		/rpt_family="Alu"	
repeat_region		complement(60338)	.60470)
repeat_region		/rpt_family="MIR"	
repeat_region		complement(61353)	.61521)
repeat_region		/rpt_family="MERS"	
repeat_region	63135. .63416		
repeat_region		/rpt_family="Alu"	
repeat_region	63724. .63889		
repeat_region		/rpt_family="MERS"	
repeat_region	64107. .64182		
repeat_region		/rpt_family="MIR"	
repeat_region	65459. .65548		
repeat_region		/rpt_family="MERS"	
repeat_region	66439. .66618		
repeat_region		/rpt_family="MER20"	
repeat_region		complement(67562)	.68084)
repeat_region		/rpt_family="MER41"	
repeat_region		complement(72861)	.73139)
repeat_region		/rpt_family="Alu"	
repeat_region	73756. .74037		
repeat_region		/rpt_family="Alu"	
repeat_region	75588. .75853		
repeat_region		/rpt_family="Alu"	
repeat_region		complement(77441)	.77729)
repeat_region		/rpt_family="Alu"	

	Query Match	Similarity	Score	DB	Length	80202;
	Best Local	99.8%;	Pred.	No. 1.2e-237;		
	Matches	1616;	Conservative	0;	Mismatches	237;
					Indels	2;
					Gaps	2;
OY	383	CCACCTCGGCTTTTAAATAGTATTTTATGATATATATTTTCCCTTTTTTTTTTTT	442			
Db	1	CCACCTCGGCTTTTAAATAGTATTTTATGATATATATTTTCCCTTTTTTTTTTTT	60			
OY	443	TAACTTTAGATGTTATGTTTCCCTATACCTTAAAGTGGTGTCTTATAGGACATATAT	502			
Db	61	TAACTTTAGATGTTATGTTTCCCTATACCTTAAAGTGGTGTCTTATAGGACATATAT	120			
OY	503	CTGGGTCTGAGATGTTATTTTATCTGATTAATCTCAACCTTTTGTGGAAGTTTAGGC	562			
Db	121	CTGGGTCTGAGATGTTATTTTATCTGATTAATCTCAACCTTTTGTGGAAGTTTAGGC	180			
OY	563	CATTACATTTAGTGTAAATTATAGACATGCTTTGATTTGCTATACCATCTTTTCATTTGT	622			
Db	181	CATTACATTTAGTGTAAATTATAGACATGCTTTGATTTGCTATACCATCTTTTCATTTGT	240			

[illegible]

623	TTTTATATGAGCCATCTTTTCACTGTCTCTTTTCACTTGTGACAAATTTCTTTAGTAC	662
241	TTTTATATGAGCCATCTTTTCACTGTCTCTTTTCACTTGTGACAAATTTCTTTAGTAC	300
689	TGAATACCTTTTTGTATTTCAATTATCTATTTGCTTTT AGTTAATACCTTTAAATTT	742
301	TGAATACCTTTTTGTATTTCAATTATCTATTTGCTTTT AGTTAATACCTTTAAATTT	360
743	TTTTTTTCTGTTTTATGTAGATTTTAAATATACATCTTTAACTTATCACAGATTACCTT	802
361	TTTTTTTCTGTTTTATGTAGATTTTAAATATACATCTTTAACTTATCACAGATTACCTT	420
803	CAATATGATTTTTCACGCTCAAGCTAATGTATGAAGACCTTATCAAGAGATATTTTCAAT	862
421	CAATATGATTTTTCACGCTCAAGCTAATGTATGAAGACCTTATCAAGAGATATTTTCAAT	480
863	TTGTGCTCTAATTTTTTATACCTA-TGTCTATATATCAATTAAGTTTGTGTGTTTGT	921
481	TTGTGCTCTAATTTTTTATACCTA-TGTCTATATATCAATTAAGTTTGTGTGTTTGT	540
922	TACCTTATTTGCTGTGTGCTGCGGTGACGAACAATTTTCTGTAAAGGCTAGATATGACAG	981
541	TACCTTATTTGCTGTGTGCTGCGGTGACGAACAATTTTCTGTAAAGGCTAGATATGACAG	600
982	GCAATCCTTGGAGATCTGTGGGTTTGTTCCATACACACACATATATACAAATATGCA	1041
601	GCAATCCTTGGAGATCTGTGGGTTTGTTCCATACACACACATATATACAAATATGCA	660
1042	GAAAGGATATCACATAATTAAGTGAAGTACACAAGTCTTTTGGCTTCCAGTGCATATATA	1101
661	GAAAGGATATCACATAATTAAGTGAAGTACACAAGTCTTTTGGCTTCCAGTGCATATATA	720
1102	AGTTTTGCTTATACCTAGCTAGTCTGTATAGTGCATATAGTTATGCTAAAAAAA	1161
721	AGTTTTGCTTATACCTAGCTAGTCTGTATAGTGCATATAGTTATGCTAAAAAAA	780
1162	CACATACCTTAATTTTAAATGCTTTATTTACTATAAAAATGCTAACATATTTAGACATT	1221
781	CACATACCTTAATTTTAAATGCTTTATTTACTATAAAAATGCTAACATATTTAGACATT	840
1222	CAGTGAATTTGTAATCTTTTTCGTGTGGAAGGCTTTCTTTATGATACATGATCGGGGG	1281
841	CAGTGAATTTGTAATCTTTTTCGTGTGGAAGGCTTTCTTTATGATACATGATCGGGGG	900
1282	TCAGGTGCTGAAGCTTAAAGGCTGTGGGAGTTTCTTAAAAACAACATGGAAGTTGCA	1341
901	TCAGGTGCTGAAGCTTAAAGGCTGTGGGAGTTTCTTAAAAACAACATGGAAGTTGCA	960
1342	TATCAGATGACTCTTCTCTTTCATGAAAGATTTTCTCTAGTGTGATGCTTTTGTATAG	1401
961	TATCAGATGACTCTTCTCTTTCATGAAAGATTTTCTCTAGTGTGATGCTTTTGTATAG	1020
1402	CATTTTATGACAGTATGAACCTTTTGAATAATTGGA-TCAATCTCTCAAAACCTGCTCT	1460
1021	CATTTTATGACAGTATGAACCTTTTGAATAATTGGA-TCAATCTCTCAAAACCTGCTCT	1080
1461	GCCTTAAACAACCTAAGTAAATATATATATCTGATCCATTTGTGTCATTTCAACAATTTT	1520
1081	GCCTTAAACAACCTAAGTAAATATATATATCTGATCCATTTGTGTCATTTCAACAATTTT	1140
1521	CACAGTGTCTTACACGAGAGTGAATTCATCTCATTTCTCTAGATGGAATCTTTGCTCAT	1580
1141	CACAGTGTCTTACACGAGAGTGAATTCATCTCATTTCTCTAGATGGAATCTTTGCTCAT	1200
1581	CCATAAGAAAGAAATCTCATCTGTCAAGTTTATATGAGATTTGCAGCAATACAGTCA	1640
1201	CCATAAGAAAGAAATCTCATCTGTCAAGTTTATATGAGATTTGCAGCAATACAGTCA	1260
1641	TGTCTTACAGGCTCATCTTCTTATATTCAGTGTCTGTGGTTTCTAACACATCTGT	1700
1261	TGTCTTACAGGCTCATCTTCTTATATTCAGTGTCTGTGGTTTCTAACACATCTGT	1320



Oy	1701	GGTTCCTTCCTCCATTTGAAGCTTGAACCTCTCCGAAGCATCCAAAGAGTGAATGCA	1766
Db	1321	GGTTCCTTCCTCCATTTGAAGCTTGAACCTCTCCGAAGCATCCAAAGAGTGAATGCA	1380
Oy	1761	CTTCTCCAAATTCCTGTAAATATTTAATTTTGAACCTCCCATGATCATGAATGTTCTT	1820
Db	1381	CTTCTCCAAATTCCTGTAAATATTTAATTTTGAACCTCCCATGATCATGAATGTTCTT	1440
Oy	1821	AATGGCACCCTGGATGTGTAATCCTTTCCAAAAGTTTTCAATTTACTTACGCCAGATCC	1880
Db	1441	AATGGCACCCTGGATGTGTAATCCTTTCCAAAAGTTTTCAATTTACTTACGCCAGATCC	1500
Oy	1881	ATCCATCCAGAGGATCCACTTTGCATATGACAGTTAATAGCTTAATGAAATGTAATTCCTCAA	1940
Db	1501	ATCCATCCAGAGGATCCACTTTGCATATGACAGTTAATAGCTTAATGAAATGTAATTCCTCAA	1560
Oy	1941	TAATTAAGCTTGAAGTTGAATTAATCTCTTGATCATCTTTCTGAAATATGATGTTGTG	2000
Db	1561	TAATTAAGCTTGAAGTTGAATTAATCTCTTGATCATCTTTCTGAAATATGATGTTGTG	1620
RESULT 8			
AF042484			
LOCUS			
DEFINITION	Homo sapiens sequence of the 80 kb core centromere protein binding region of the mar del (10) neocentromere corresponding to human chromosome 10q25.2, complete sequence.		
ACCESSION	AF042484		
VERSION	AF042484.1		
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 80155)		
TITLE	Barry,A.E., Howman,E.V., Cancelli,M.R., Saffery,R. and Choo,K.H.		
JOURNAL	Sequence analysis of an 80 kb human neocentromere		
PUBMED	Hum. Mol. Genet. 8 (2), 217-227 (1999)		
REFERENCE	9931329		
AUTHORS	2 (bases 1 to 80155)		
TITLE	Barry,A.E., Howman,E.V., Cancelli,M.R., Saffery,R. and Choo,A.		
JOURNAL	Direct Submission		
FEATURES	Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville, Melbourne, Victoria 3052, Australia		
SOURCE	Location/Qualifiers		
	1. 80155		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="10"		
	/map="10q25.2"		
	/note="Sequence information obtained from BAC clones prepared from cell line BE201-18-5f"		
	complement(join(49..525,3628..4545,4849..5053))		
	/rpt_family="TJMEC"		
	/rpt_type="dispersed		
	join(560..597,3184..3281)		
	/rpt_family="HERS8"		
	/rpt_type="dispersed		
	complement(join(598..2553,2716..3183))		
	/rpt_family="TJGGR1"		
	/rpt_type="dispersed		
	2558..2581		
	/standard_name="T1"		
	/note="100% homology"		
	/rpt_type="direct		
	/rpt_unit="(caggcacagctg)2"		
	2570..2715		
	/rpt_family="Alu/g.x"		
	/rpt_type="dispersed		
	3021..3050		
	repeat_region		

```

repeat__region
/standard_name="r3"
/note="83.3% homology"
/rpt_type=direct
/rpt_unit="(taacaagtgc)3"
3283..3558
/rpt_family="AlusX"
/rpt_type=dispersed
4173..4211
/standard_name="r3"
/note="82.1% homology"
/rpt_type=direct
/rpt_unit="(aaaaaatatttc)3"
complement(4546..4648)
/rpt_family="AlusX"
/rpt_type=dispersed
complement(5909..6208)
/rpt_family="Alusg"
/rpt_type=dispersed
join(6232..6573,6638..6688)
/rpt_family="MLT2CB"
/rpt_type=dispersed
6598..6645
/standard_name="AC24"
/note="100% homology"
/rpt_type=tandem
/rpt_unit="ac"
6598..6637
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="ca"
7091..7134
/rpt_family="MIR"
/rpt_type=dispersed
7808..7911
/rpt_family="L2"
/rpt_type=dispersed
8349..8524
/rpt_family="MIR"
/rpt_type=dispersed
complement(8839..9124)
/rpt_family="Aluub"
/rpt_type=dispersed
9180..9280
/rpt_family="MIR"
/rpt_type=dispersed
10070..10170
/note="low complexity"
/rpt_family="purine-rich"
10097..10136
/standard_name="GA1"
/note="82.5% homology"
/rpt_type=direct
/rpt_unit="(gaagaaggg)4"
10175..10339
/rpt_family="FRAM"
/rpt_type=dispersed
complement(10899..11197)
/rpt_family="AlusX"
/rpt_type=dispersed
join(11827..11757,12434..12512,13443..13562)
/note="98% identity"
/db_xref="dbEST:AA74571"
11956..12028
/standard_name="AT4"
/note="100% homology"
/rpt_type=tandem
/rpt_unit="(attc)8"
complement(12010..12312)
/rpt_family="Aluub"
/rpt_type=dispersed
12327..12433
/rpt_family="L2"
/rpt_type=dispersed

```

```
repeat_region complement(13946..14235)
/rpc_family="AluSg"
/rpc_type=dispersed
repeat_region 15130..15159
/note="low complexity"
/rpc_family="Alu-rich"
repeat_region 15164..15753
/standard_name="AT28"
/note="82.1% homology"
/rpc_type=direct
/rpc_unit="(atgataatgctatagacataatc)21"
join(15170..15289,15410..15612)
/note="microsatellite"
/rpc_type=tandem
/rpc_unit="ca"
repeat_region 15617..15731
/note="microsatellite"
/rpc_type=tandem
/rpc_unit="cata"
repeat_region 16004..16100
/rpc_family="L2"
/rpc_type=dispersed
repeat_region complement(16128..16364)
/rpc_family="L2"
/rpc_type=dispersed
repeat_region complement(16407..16592)
/rpc_family="L2"
/rpc_type=dispersed
repeat_region complement(16657..16889)
/rpc_family="MIR"
/rpc_type=dispersed
repeat_region 16914..16936
/note="low complexity"
/rpc_family="Alu-rich"
repeat_region 17102..17386
/rpc_family="AluSx"
/rpc_type=dispersed
repeat_region 17406..17564
/standard_name="GA2"
/note="81.8% homology"
/rpc_type=direct
repeat_region 17416..17543
/note="microsatellite"
/rpc_type=tandem
/rpc_unit="gaaa"
repeat_region 17647..17702
/rpc_family="MIR"
/rpc_type=dispersed
repeat_region complement(19863..20161)
/rpc_family="AluSx"
/rpc_type=dispersed
repeat_region complement(20847..20913)
/rpc_family="L2"
/rpc_type=dispersed
repeat_region complement(21366..21414)
/note="microsatellite"
/rpc_type=tandem
/rpc_unit="gaaa"
repeat_region 21375..21639
/standard_name="CT1"
/note="75% homology"
/rpc_type=dispersed
repeat_region complement(21415..21536)
/note="microsatellite"
/rpc_type=tandem
/rpc_unit="cgaaa"
repeat_region complement(21524..21646)
/note="low complexity"
/rpc_family="purine-rich"
repeat_region 22403..22646
/rpc_family="L1M4"
/rpc_type=dispersed
repeat_region 22560..22604
```

```
Query Match 78.8%; Score 1575.2; DB 8; Length 80155;
Best Local Similarity 99.1%; Fred.No. 6.1e-235;
Matches 1605; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

/standard_name="T4"
/note="80% homology"
/rpc_type=direct
/rpc_unit="(aataacataatc)3"

383 CCACCTGGCTTTTAAATAATAGTATTTTATGTAATATTTTCTTTTATTTT 442
Db 1 CCAATTTGGGTTTTTAAATAATAGCAATTTTATGGAAAAATTTTCCCTTTT 60
443 TAAAGTTTAAATGTTATGTTTCTTTTAACTTAAAGTGGTCTTAAAGGACATAT 502
Db 61 TAAAGTTTAAATGTTATGTTTCTTTTAACTTAAAGTGGTCTTAAAGGACATAT 120
503 CTGGGCTTGAATGTTATGTTTAACTGTAATCTCAACCTTTTGTGAGGTTAGGC 562
Db 121 CTGGGCTTGAATGTTATGTTTAACTGTAATCTCAACCTTTTGTGAGGTTAGGC 180
563 CATTTACATTTAAGTAAATTAAGACATGTTTGAATTTGCTAATACCATTTTCA 622
Db 181 CATTTACATTTAAGTAAATTAAGACATGTTTGAATTTGCTAATACCATTTTCA 240
623 TTTATATGTAAGCCATCTTTTCAATGTTTTCATCTTTGACCATTTTCTTATGAC 682
Db 241 TTTATATGTAAGCCATCTTTTCAATGTTTTCATCTTTGACCATTTTCTTATGAC 300
683 TGAATACCTTTTTCATTTATTTATATATCTATGCTTTTATGATATCTTTAAAT 742
Db 301 TGAATACCTTTTTCATTTATTTATATATCTATGCTTTTATGATATCTTTAAAT 360
743 TTTTTCCTGTTTATGATGATTTATATATATACCTTTTATGACATTAACCTT 802
Db 361 TTTTTCCTGTTTATGATGATTTATATATATACCTTTTATGACATTAACCTT 420
803 CAATAGATATTTTACCACTCAAGTAAATGTAAGAACTTCAAGAGTATATTTTCA 862
Db 421 CAATAGATATTTTACCACTCAAGTAAATGTAAGAACTTCAAGAGTATATTTTCA 480
863 TCTGTCCTCAATTTTATATGCTA-TGCTAATATACATTAAGTTTGTGTTGTTT 921
Db 481 TCTGTCCTCAATTTTATATGCTAATATATATATTAATTAAGTTTGTGTTGTTT 540
922 TACCTTATGCTGTTGGCTGGGCTGAGCAACATTTCTGTAAGGAGCTAGATGAC 981
Db 541 TACCTTATGCTGTTGGCTGGGCTGAGCAACATTTCTGTAAGGAGCTAGATGAC 600
982 GCATACCTTGAAGATACGTGTGGGTTGGTTCATACCAACACATATATCAAAATG 1041
Db 601 GCATACCTTGAAGATACGTGTGGGTTGGTTCATACCAACACATATATCAAAATG 660
1042 GAAGTGAATATCACAATTAAGAGTCAACAAGCTTTGGCTTCCAGTGCATATTA 1101
Db 661 GAAGTGAATATCACAATTAAGAGTCAACAAGCTTTGGCTTCCAGTGCATATTA 720
1102 AGTTTGTCTTACTACACTGATGCTGTTAAGTGCATATGTTATGCTTAAAAAA 1161
Db 721 AGTTTGTCTTACTACACTGATGCTGTTAAGTGCATATGTTATGCTTAAAAAA 780
1162 CACATACCTTAAATTTTAAATGCTTATTAATAAAATGCTAACAATGTTGACAT 1221
Db 781 CACATACCTTAAATTTTAAATGCTTATTAATAAAATGCTAACAATGTTGACAT 840
1222 CAGTGAATGTAATCTTTTGTGCTGAGAGGCTTTTCTTATATGATGCTATGGGG 1281
Db 841 CAGTGAATGTAATCTTTTGTGCTGAGAGGCTTTTCTTATATGATGCTATGGGG 900
1282 TCAGGTCGTAAGCTTAAAGGTCGTGTCAGTTCTTAAACAACAGTAAGATGCA 1341
Db 901 TCAGGTCGTAAGCTTAAAGGTCGTGTCAGTTCTTAAACAACAGTAAGATGCA 960
```

QY	1342	TATCGATTGACGCTCTTCCTTTATGAAAGATTTCTCTAGGTGGAGTGCCTTTGGATG	1401
Db	961	TATCGATTGACGCTCTTCCTTTATGAAAGATTTCTCTAGGTGGAGTGCCTTTGGATG	1020
QY	1402	CATTTTATGACACAGTAGAATCTCTTGGAAATTGGA-TCAATCCCTCAACCCGTCTCT	1460
Db	1021	CATTTTATGACACAGTAGAATCTCTTGGAAATTGGAAGTCATCTCTCAACCCGTCTCT	1080
QY	1461	GCTTTTAACACCTAAGTTAATATAATATTCGAAATCCATGTGTGCATTTCAACAATTTT	1520
Db	1081	GCTTTTAACACCTAAGTTAATATAATATTCGAAATCCATGTGTGCATTTCAACAATTTT	1140
QY	1521	CACAGTGTCTTACACAGAGTAGATTCATCTCATTTCTCTGAGATGAGATCTTGTCTCAT	1580
Db	1141	CACAGTGTCTTACACAGAGTAGATTCATCTCATTTCTCTGAGATGAGATCTTGTCTCAT	1200
QY	1581	CCATTAAGAAGAAATTCCTCATCTGTTCGAAGTTTATCATGAGATTCGACGAATCAGTCA	1640
Db	1201	CCATTAAGAAGAAATTCCTCATCTGTTCGAAGTTTATCATGAGATTCGACGAATCAGTCA	1260
QY	1641	TGTCTTCAGGCTCACTTCACCTTTAATTCAGTTCCTGCTGTCTTACACATCTGT	1700
Db	1261	TGTCTTCAGGCTCACTTCACCTTTAATTCAGTTCCTGCTGTCTTACACATCTGT	1320
QY	1701	GCTTCCTTCCTTCATTTGAAGTCTTGAACTCTCCAAAGTCATTCATGAGGGTTGGAATCGA	1760
Db	1321	GCTTCCTTCCTTCATTTGAAGTCTTGAACTCTCCAAAGTCATTCATGAGGGCTGGAATCGA	1380
QY	1761	CTTCTTCGAAATTCCTGTTAATTTATTTATTTGACCTCCCATGAAATGATGTTCTT	1820
Db	1381	CTTCTTCGAAATTCCTGTTAATTTATTTATTTGACCTCCCATGAAATGATGTTCTT	1440
QY	1821	AATGGACACTGGAATGTGAATCCCTTCCAAAAGGTTTCAATTACTTACTTCAGATCC	1880
Db	1441	AATGGACACTGGAATGTGAATCCCTTCCAAAAGGTTTCAATTACTTACTTCAGATCC	1500
QY	1881	ATCCATCCAGAGATGCATCTTCAATGCGAGTTATAGCCTTATAGGAATGATTTCTTCAA	1940
Db	1501	ATCCATCCAGAGATGCATCTTCAATGCGAGTTATAGCCTTATAGGAATGATTTCTTCAA	1560
QY	1941	TAAATAGGCTTGAAGTTGAATTAATCTCCTTGATTCATTTTCTGCAAAATAGATGTGTG	2000
Db	1561	TAAATAGGCTTGAAGTTGAATTAATCTCCTTGATTCATTTTCTGCAAAATAGATGTGTG	1620
RESULT 9			
LOCUS	AR612649/c	2418 bp	DNA linear PAT 15-DEC-2004
DEFINITION	Sequence 148 from patent US 6828097.		
ACCESSION	AR612649		
VERSION	AR612649.1 GI:56668471		
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2418)		
TITLE	Knoll, J.H.M. and Rogan, P.K.		
JOURNAL	Single copy genomic hybridization probes and method of generating		
FEATURES	Name Patent: US 6828097-A 148 07-DEC-2004;		
source	The Childrens Mercy Hospital; Kansas City, MO		
	Location/Qualifiers		
	1..2418		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Query Match	31.8%;	Score 636.8;	DB 6; Length 2418;
Best Local Similarity	82.1%;	Pred. No. 2.6e-89;	
Matches 859; Conservative	1;	Mismatches 133;	Indels 53; Gaps 9;
QY	979	CAGGATACCTTGGAGATACCTGTGGGTTGGTTGCATACCAACCAATATATG	1038

Db	2418	CAGGCATACCTCGGAGATATTGGGGGTGGGTTCCAGACCAACGCGAATA-----	2370
QY	1039	CAAGAAGTGATATCAOAAATPAAAGTGACACACAGCTTTTGGCTTCCAGTCATAT	1098
Db	2369	----AAGGAATATCGCAATPAAAGCGAGTCACACGAATTTTTTGGTTCCAGTCATAT	2314
QY	1099	AAAAGTTTGGTTATACTACACGTGATCTGTAAAGTGCATAPAGTTATATGCTPAAA	1158
Db	2313	AAAAGTTATGTTAAGGCTATACGTGATCTGTAAAGTGCATAPAGTTATATGCTPAAA	2254
QY	1159	AA-----ACACATACCTTAATTTTAAAATVGTCTTATTAATAAAAATCTCAACAATCATTT	1214
Db	2253	AACATATGATACCTTAATTTTAAAATACTTTTATTTCTTAAAAATATCTAACATCATCT	2194
QY	1215	GAGCATTCAGTAGATTTGATTCCTTTTGTCTGTGGAAAGGCTTTTCTTATATGATGACGA	1274
Db	2193	GAGCTTCAGGAGATCGTAATCTTTTGTGTGTGGAGAGGCTTGCTGATATGTTGATGGC	2134
QY	1275	T-----CGGGGTCAGGTGCTGAAGCTTAAGGTGGCTGTGGCAGTTTCTTAAA-	1322
Db	2133	TGCTGACTGATCAGAGGTGTGTGTGTGCGAAGGTGGGTGGCTGTGCAATTTCTTAAAA	2074
QY	1323	----ACAAAGTAGAAGATTCGAATATACAGTTGACTCTTCCTTTCATGAAGAATTTCTCTC	1378
Db	2073	TAAACAACAATGAAAGTTTGGCCGATCAATTTGACTCTTCCTTTCATGAAGAATTTCTCTG	2014
QY	1379	TAGTGTGTGATGCTTTTGTGATAGCATTTTATGCAAGTAGAACTTCTTGAATAATGGA-	1437
Db	2013	TAGATATGTGATGCTTTTGTGATAGCATTTTATGCAAGTAGAACTTCTTGAATAATGGAAG	1954
QY	1438	TCAAATCCTCTCAAAACCTGCTCTGCTTTAAACAACCTTAAGTTAATATATATTTCTGAATCC	1497
Db	1953	TCAAATCCTCTCAAAACCTGCGCTGCTTTATCAACTAAGTTTATGTATATTTCTAAATCC	1894
QY	1498	ATTGTGTTCATTTCAACAATTTTCAAGTGTCTTACACAGAGTAGAATTCATCTCATTT	1557
Db	1893	TTTGTGTTCATTTCAACAATTTTCAAGATCTTTCACAGAGTAGAATTCATCTCAAGA	1834
QY	1558	CCTGAGATGGAATCTTGTGCTCATCTCATPAAAGAAATTCCTCATCTGTTCAGAGTTTATCT	1617
Db	1833	AACCACTT-----TCTTTGCTCATCTCATPAAAGAGCACTCTCATCTGTTCAAGTTTATCT	1778
QY	1618	ATGAGATTTGACAGCAATACAGTCATGTCTTCAGGCTTCACCTTCACTTTAATTCAGTCTCT	1677
Db	1777	ATGAGATTTGACAGCAATACAGTCATGTCTTCAGG-----CTCACTTCTAATTCAGTCTCT	1723
QY	1678	CTTGTCTGTTCTTACCAATCTGTGTGTTCTTCTCCATTTGAAGTTTGAACCTCTCCAAAG	1737
Db	1722	CTTGTCTTATTTCCACCAATCTGTGAGTTAATCTCTCCACTGAATCTTTAAACCCCTCCAAAG	1663
QY	1738	TCATCCATGAGGGTTGGATCGAATCTTCCCAAAATTCCTGTATATTTATTTATTTGA--	1795
Db	1662	TCATCCATGAGGGTTGGATCGAATCTTCTCCAAATCTCTGTATATTTGAATTTTGAACC	1603
QY	1796	-CCTCCCATGAAATCATGAATGTTCTTAAATGGCAACTGGAAATGTGTGAATCCTTTCCAAAG	1854
Db	1602	TCTCCCATGAAATCATGAATGTTCTTAAATGGCAATGTGTGAATCCTTTCCGAAGAG	1543
QY	1855	GTTTTCATTTACTTAAGTCCAGATCCATCCATCCAGAGATCCACCTTTCATATCCAGTTA	1914
Db	1542	GTTTTCATTTACTTAAGTCCAGATCCAT-----CAGAGGAATCACTATCATATGGCAGCTA	1488
QY	1915	TAGCTTATGGAATGTATTTCTTCAATPAAATPAAAGCTTGAAGTTGAAATTTACTCTTGAT	1974
Db	1487	TAGCTTATPAAATGTATTTCTTCAATPAAATPAAAGCTTGAAGTTGAAATTTACTCTTGAT	1428
QY	1975	CCATTTTCTGCAAAATAGATGTTGTG	2000
Db	1427	CCATGGGCTGCAGAAATGATGTTGTG	1402

LOCUS HSU49973 2418 bp DNA linear PRI 08-JUN-2005  
DEFINITION Human Tigger1 transposable element, complete consensus sequence.  
ACCESSION U49973  
VERSION U49973.1 GI:2226003  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2418)  
AUTHORS Smit,A.F. and Riggs,A.D.  
TITLE Tiggers and DNA transposon fossils in the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (4), 1443-1448 (1996)  
PUBMED 8643651  
REFERENCE 2 (bases 1 to 2418)  
AUTHORS Robertson,H.M.  
TITLE Members of the pogo superfamily of DNA-mediated transposons in the human genome  
JOURNAL Mol. Gen. Genet. 252 (6), 761-766 (1996)  
PUBMED 8917322  
REFERENCE 3 (bases 1 to 2418)  
AUTHORS Robertson,H.M.  
TITLE Direct Submission  
JOURNAL Submitted (15-FEB-1996) Hugh M. Robertson, Entomology, University of Illinois at Urbana-Champaign, 505 S. Goodwin, Urbana, IL 61801, USA  
COMMENT On Jun 28, 1997 this sequence version replaced gi:1224064.  
FEATURES  
source Location/Qualifiers  
1..2418  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
1..2418  
/note="consensus sequence based on 50 full-length genomic sequences"  
repeat\_region  
/transposon="Tigger1"  
1..13  
/rpt\_type=inverted  
425..1789  
/note="ORF1, MER37, similar to pogo element"  
/citation=[1]  
/codon\_start=1  
/product="putative transposase"  
/protein\_id="AAB61714.1"  
/db\_xref="GI:2226004"  
/translation="MAKSSSRKSRSTSLTNOKLEMTKLSRSGSKAETGKGLIR  
QTVSVVNAKEKFLKEIKATPVNTRMTIKRNSLIADMEKVLVNI EDQTSNIPUSQ  
SLIQKALTLFNSMAERGEAEAEKLEKSRGMFMRFKRSILHNI KVQGEASADGE  
AAASYPEDIKIIDEGGYKQOI FNVDETAFAWKKPSRTFLAREBSKMPFKASDR  
LTLIGANAAGDFKLPMLIYSENPBALKNYKSTLPVLYKNNKAMWTALHPTAMF  
TEYKPVETVYCEKKISPKILLIDNAGHPALMEWKINIVMPMANTSLIOPM  
DGVISTKRSYLIKNTFRKATIAIDSDSGGSKLTKFWKGFITLDAIKNIRSWE  
EVKISTLTGVWKKLIPITLMDPEGFKTSVEVADVETARELELEVEBEDVETELDS  
HDKT"  
1811..2206  
/note="ORF2: function unknown"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAB61715.1"  
/db\_xref="GI:2226005"  
/translation="MDEQRKFLMESTPGEADVNI VENTTKOLEYIILVYKAAAGF  
BRIDNFERSTVVGKMLNSISACVREI FFERBSQSMRQSLISYFKLQPPQPSATV  
TIIQQPSTSRDPPAKRLRLAEGSDDR"  
2218..2223  
2405..2418  
/rpt\_type=inverted  
ORIGIN  
polyA\_signal  
repeat\_region  
Query Match 31.7%; Score 634; DB 8; Length 2418;  
Best Local Similarity 82.0%; Pred. No. 7,1e-89;  
Matches 858; Conservative 0; Mismatches 135; Indels 53; Gaps 9;

QY 979 CAGGATACCTTGGAGATACCTGTGGTTGGTTGCCATACCCACCAATATATACAAATATG 1038  
Db 2418 CAGGATACCTTGGAGATATTTGGGTTCCGATCCAGACCCGCAATTAAGCA----- 2364  
QY 1039 CAAGAAGTGAATATACAAATAAGTACACAAAGCTTTGGCTCCAGTGCATAT 1098  
Db 2363 -----ATATGCATTAAGGAGATCACAGAAATTTTGGTTTCCAGTGCATAT 2314  
QY 1099 AAAAGTTTGGTTATATACATACAGTATGCTGTATAGTGCATATAGTATGTATAAA 1158  
Db 2313 AAAAGTTATGTATACATATACGTATGCTATTAAGTGTGCAATGATATATGTATAAA 2254  
QY 1159 AA----AACAATCTTAAATTTTAAATGCTTATATCTTAAATAAATGCTAACATCATTT 1214  
Db 2253 AACAATGTACATACCTTAAATTTTAAATACTTATTTGTAAATAAATGCTAACATCATCT 2194  
QY 1215 GAGCATTCAGAGATGTATATCTTTTCTGATGAGAGTCTTTCTTATATGATGATCTGA 1274  
Db 2193 GAGCTTCAGAGATGCTATATCTTTTCTGATGAGAGTCTTTGCTGATGATGATGATGCT 2134  
QY 1275 T-----CGGAGGTCAGAGTCTGAGAGCTTAGAGTGTGCTGTGCAATTTCTTAA- 1322  
Db 2133 TCTGATGATGATGAGAGTGTGATGCTGAGAGTGTGCTGTGCAATTTCTTAA- 2074  
QY 1323 ----ACAACATGAAATTTGCAATATCAATGATCTCTTCAATGAAAGTTTCTGTC 1378  
Db 2073 TAAACAACAATGAAATTTGCGATGATGATGATGATGATGATGATGATGATGATGATG 2014  
QY 1379 TAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1437  
Db 2013 TAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1954  
QY 1438 TCAATCTCTCAAACT 1497  
Db 1953 TCAATCTCTCAAACT 1894  
QY 1498 ATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1557  
Db 1893 TTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1834  
QY 1558 CCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1617  
Db 1833 AACCACTT---TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1778  
QY 1618 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677  
Db 1777 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1723  
QY 1678 CTGTGCTGTTTACACATCTGTGTTCTTCTCTCAATTTGAAGTCTTGAACCTTCCAA 1737  
Db 1722 CTGTGCTGTTTACACATCTGTGTTCTTCTCTCAATTTGAAGTCTTGAACCTTCCAA 1663  
QY 1738 TCATCAGAGGTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795  
Db 1662 TCATCAGAGGTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1603  
QY 1796 -CCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1854  
Db 1602 TCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1543  
QY 1855 GTTTTCATTTTACCTAGTCCAGATCCATCCAGAGATCCATCTTCAATGACGTTA 1914  
Db 1542 GTTTTCATTTTACCTAGTCCAGATCCAT-----CAGAGGAATCAGATCTATGAGCGCTA 1488  
QY 1915 TAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1974  
Db 1487 TAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1428  
QY 1975 CCATTTTCGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2000  
Db 1427 CCATGGGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1402

RESULT 11  
AC092165 100020 bp DNA linear PRI 30-APR-2005  
LOCUS Homo sapiens BAC clone RP11-753F4 from 2, complete sequence.  
DEFINITION AC092165  
AC092165.4 GI:18056750  
VERSION HTG.  
KEYWORDS Homo sapiens (human)  
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 100020)  
Haglund, K. and Abbott, S.  
The sequence of Homo sapiens BAC clone RP11-753F4  
Unpublished (2001)  
2 (bases 1 to 100020)  
Waterston, R.H.  
Direct Submission  
Submitted (22-JUN-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 100020)  
Waterston, R.H.  
Direct Submission  
Submitted (04-JAN-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 100020)  
Waterston, R.  
Direct Submission  
Submitted (21-FEB-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
5 (bases 1 to 100020)  
Wilson, R.K.  
Direct Submission  
Submitted (30-APR-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jan 4, 2002 this sequence version replaced gi:17981677.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Web site: http://genome.wustl.edu  
Contact: submissions@wustl.wustl.edu  
----- Summary Statistics  
Center project name: H\_NH0753F04  
-----  
COMMENT  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.  
-----  
MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
http://genome.wustl.edu  
-----  
SOURCE INFORMATION:  
The RPC1-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,  
Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
and coworkers at http://www.chori.org

VECTOR: pBAC3.6  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-762N20, 2000 bp overlap;  
the clone sequenced to the right is RP11-485D7. Actual start of  
this clone is at base position 98039 of RP11-762N20; actual end is  
at base position 100020 of RP11-753F4.  
Data from AC09338 and AC074188 was used to finish this clone,  
AC092165. Polymorphisms have been identified between AC074188 and  
AC092165.  
FEATURES  
source  
Location/Qualifiers  
1..100020  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-753F4"  
/clone\_1fb="RPC1-11"  
complement(8894..16888)  
/gene="ECE1"  
complement(join(8894..9319,9466..9542,9785..9880,  
10158..10223,10573..10697,10849..10916,11485..11536,  
11617..11675,11918..12021,12172..12246,12483..12581,  
13068..13290,13539..13663,13868..13960,14048..14158,  
14266..14334,14935..15821,16786..16888))  
/gene="ECE1"  
complement(join(9220..9319,9466..9542,9785..9880,  
10158..10223,10573..10697,10849..10916,11485..11536,  
11617..11675,11918..12021,12172..12246,12483..12581,  
13068..13290,13539..13663,13868..13960,14048..14158,  
14266..14334,14935..15720))  
/gene="ECE1"  
/note="Homo sapiens endochelin converting enzyme-like 1  
(ECE1)", mRNA.; H\_NH0753F04.1  
This gene was based on gi(4758231)"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AA124101.1"  
/db\_xref="GI:62988714"  
/translation="MEPPSYSLTAHYDERQEVKYSRCGAGAGASLPFGPLGAARS  
ATGARGSLPRMWRREVCLSGVFAAGCALIAMLALKYLPVAGGACGEGCER  
KAFARARFPIANLNASIDPCDFSPACGWLRRHAIIPDDLT/VTGTLAIGEONER  
LRRLARPGCGGGAORKVRAFPFSCIDMRERIEIIRPMLLEVIDCGMPLGABE  
RPGVAARDINRLYKAGCVSYAALFSLTUSLDNRSSRYIRIDDSGLTPERTLY  
LAQDSESKTILAAVRFEMERVLISLGAADVEQKAEIILQVEQLANITVSEHDLRD  
VSMYNKYTLGQLKTPHLRKMLLDQIFQDFSEBEVVVLADVMQVSQLIRST  
PFRVLAHTLVWRVVVLSLSPPREALHELQMEGSDKQDELARVCLAGANHHV  
MALGALFVHEHFSASAKAKVQVLDVRYTILGQRLBELDDMDAETRAARALQYMMV  
MVGYPDPLIKDAVDKEYEFVEHYETKRNITNSIRFISQSVKKIROYDSSMLP  
FDALNAVYLPNKMVPFAGILQPTLPDPPQSLNAGIGITITIGHELTGHDWGGQ  
YDLSGNLLHMTWASYSRFLRAECIVLYNDFTVYNRVNKGHTLGENIADMGGLK  
HYAVQKWRSEHPEHPPLRLKTYTDQDFLFAQNCIKRSOSIYIQLVLTDKHAPB  
HYRVGVSVQSEFEFRAPHCPSDPMNAHKCSVM"  
14636..17331  
/note="CpG island (GC=72.6, o/e=0.83, #CpGs=313)"  
31824..32934  
/note="CpG island (GC=68.7, o/e=0.83, #CpGs=95)"  
50972..53518  
/note="CpG island (GC=70.7, o/e=0.66, #CpGs=204)"  
55279..64562  
/gene="CHRD"  
join(55279..55334,55596..55741,56468..56512,57329..57438,  
57568..57723,57929..58038,59006..59206,60419..60530,  
60609..60723,62998..63202,63291..63409,64197..64562)  
/gene="CHRD"  
join(55283..55334,55596..55741,56468..56512,57329..57438,  
57568..57723,57929..58038,59006..59206,60419..60530,  
60609..60723,62998..63202,63291..63409,64197..64562)  
/gene="CHRD"  
/note="Homo sapiens cholinergic receptor, nicotinic, delta  
polypeptide (CHRD)", mRNA.; H\_NH0753F04.2  
This gene was based on gi(4557460)"



TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
 David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
 Dugan-Kocha, S., Durbin, K.J., Fernandez, C., Ferriquto, D.,  
 Forcum-Tamsey, J., Frantz, P., Ganes, R., Gorrell, J. H., Gorrell, L. L.,  
 Guetara, M., Harris, K., Hernandez, J., Hodgson, A., Hogues, M.,  
 Holloway, C., Hosak, H., Jackson, L. E., Jackson, L., Jia, Y., Jones, M.,  
 Kelly, S., Kondewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
 Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozado, R. J., Lu, J.,  
 Lucier, R., Martin, R., Martinez, C., McLeod, M. P., Mei, G., Morgan, M.,  
 Morris, S., Naeh, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,  
 Oswald, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L. L.,  
 Qules, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,  
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sungarc, R.,  
 Tabori, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wabnah, M.,  
 Watlington, S., Weinstein, G., Weinstein, I. R., Williamson, A.,  
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and  
 Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 158276)  
 Worley, K. C.  
 Direct Submission  
 Submitted (19-Oct-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Sep 4, 2000 this sequence version replaced gi:9929521.

```

Center Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: HMOR
Center clone name: RP11-91N19
-----
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 135987 bases at least Q40
Consensus quality: 145984 bases at least Q30
Consensus quality: 149762 bases at least Q20
Estimated insert size: 149944; sum-of-contigs estimation
Estimated insert size: 249006; agarose-gel estimation
Quality coverage: 2.8x in Q20 bases; agarose-gel estimation
Quality coverage: 4.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a "working draft" sequence. It currently
  consists of 15 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.

1
44367: contig of 44367 bp in length
44368 44467: gap of unknown length
44468 73078: contig of 28611 bp in length
73079 73178: gap of unknown length
73179 94206: contig of 21028 bp in length
94207 94306: gap of unknown length
94307 112786: contig of 18482 bp in length
112789 112888: gap of unknown length
112889 120323: contig of 7435 bp in length
120324 120423: gap of unknown length
120424 130440: contig of 10017 bp in length
130441 130540: gap of unknown length
130541 135818: contig of 5278 bp in length
135819 135918: gap of unknown length
135919 141234: contig of 5316 bp in length
141235 141334: gap of unknown length
141335 145232: contig of 3898 bp in length
145233 145332: gap of unknown length
145333 149660: contig of 4328 bp in length

```

	*	149661	149760:	gap of unknown length	
	*	149761	150931:	contig of 1171 bp in length	
	*	150932	151031:	gap of unknown length	
	*	151032	152757:	contig of 1726 bp in length	
	*	152758	152857:	gap of unknown length	
	*	152858	154933:	contig of 2076 bp in length	
	*	154934	155033:	gap of unknown length	
	*	155034	156717:	contig of 1664 bp in length	
	*	156718	156817:	gap of unknown length	
	*	156818	158276:	contig of 1459 bp in length.	
FEATURES					
SOURCE					
			1..158276		
			/organism="Homo sapiens"		
			/mol_type="Genomic DNA"		
			/db_xref="taxon:9606"		
			/clone="RP11-91N19"		
	gap		44368..44467		
			/estimated_length=unknown		
	gap		73079..73178		
			/estimated_length=unknown		
	gap		94207..94306		
			/estimated_length=unknown		
	gap		112789..112888		
			/estimated_length=unknown		
	gap		120324..120423		
			/estimated_length=unknown		
	gap		130441..130540		
			/estimated_length=unknown		
	gap		135819..135918		
			/estimated_length=unknown		
	gap		141235..141334		
			/estimated_length=unknown		
	gap		145233..145332		
			/estimated_length=unknown		
	gap		149661..149760		
			/estimated_length=unknown		
	gap		150932..151031		
			/estimated_length=unknown		
	gap		152758..152857		
			/estimated_length=unknown		
	gap		154934..155033		
			/estimated_length=unknown		
	gap		156718..156817		
			/estimated_length=unknown		
ORIGIN					
Query Match		26.3%;	Score 526.6;	DB 14;	Length 158276;
Best Local Similarity		78.7%;	Pred. No. 66-73;		
Matches 767;		Conservative	0;	MisMatches 169;	Indels 39; Gaps 10;
Gy		1051	ATCAACAATAAGTAGTGACACA--CAAGTCCTTTGGGCTCCACAGCATATATAAAGTTTGC		1109
Db		109769	AACACAAAGAACAGACGTACATGAATGTTTTTGGTTCCCAATGCATATATAAAGTAATGT		109828
Gy		1110	TTAATACTACCTGTAGTCGTGTAAGTGTGCATAG---TGTATGTCTAAAAAACAACAT		1166
Db		109829	TTACCTATACCTGTAGACACAGCAAGAGCATAGATTGTCTAATATAAACAATATATACAT		109886
Gy		1167	ACCTTAATTTAAAATGCTTTATTACTATAAAAATGCTAACATC-ATTGAGCATTCAGT		1225
Db		109889	ACCTAAATTTAAAATATCTTTATATGCTAAAAATGCTGATTTATCATCTGAGGCTTCGGT		109948
Gy		1226	GAGTTGTATCTTTTGTGTGGTGAAGAAGTCTTTCTTATGTATGACTGAT-----		1275
Db		109949	GAGTGTATCTTTTGTGTGGTGAAGAAGTCTTCCTCGAAGTGTGATGTTGCTGACTGGT		110008
Gy		1276	-CGGGGGGTGAGTGTCAAAGCTTAAGGTGAGTGTGGAATTTCTTAA-----ACAAAG		1329
Db		110009	CAGGGTGTGCTGCTGTAAGGTGAGGTGAGTGTGGAATTTCTTAAATTAAGAATTCGG		110068
Gy		1330	TGAAGATTGCATATACAGTTGACTTCTCTTTCATGAAAAGATTTCTCTTACTAGTGTGAT		1389
Db		110069	TGAAGCTTTTGCATGAGATTGACTTCTCTTTCATGAAAAGATTTCTCTGTACATGCCAT		110128



```

Webster code: M05SC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0341G20
----- Summary Statistics -----
Sequencing vector: M13; 94%
Sequencing vector: plasmid; 6%
Chemistry: Dye-terminator ET; 94% of reads
Chemistry: Dye-terminator Big Dye; 6% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 167166 bases at least Q40
Consensus quality: 170689 bases at least Q30
Consensus quality: 172797 bases at least Q20
Insert size: 186000; agarose-fp
Insert size: 178216; sum-of-contigs
Quality coverage: 5.07 in Q20 bases; agarose-fp
Quality coverage: 5.39 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
1 3838: contig of 3838 bp in length
*
* 3839 3938: gap of unknown length
*
* 3939 9367: contig of 5429 bp in length
*
* 9368 9467: gap of unknown length
*
* 9468 15261: contig of 5794 bp in length
*
* 15262 15361: gap of unknown length
*
* 15362 21615: contig of 6254 bp in length
*
* 21616 21715: gap of unknown length
*
* 21715 28467: contig of 6752 bp in length
*
* 28468 28567: gap of unknown length
*
* 28568 36770: contig of 8203 bp in length
*
* 36771 36870: gap of unknown length
*
* 36871 48844: contig of 12074 bp in length
*
* 48844 49044: gap of unknown length
*
* 49045 49044: gap of unknown length
*
* 49045 60729: contig of 11685 bp in length
*
* 60730 60829: gap of unknown length
*
* 60830 74071: contig of 13242 bp in length
*
* 74072 74171: gap of unknown length
*
* 74172 86808: contig of 11637 bp in length
*
* 86809 86808: gap of unknown length
*
* 86809 101607: contig of 15699 bp in length
*
* 101608 101707: gap of unknown length
*
* 101708 120321: contig of 18514 bp in length
*
* 120322 120321: gap of unknown length
*
* 120322 144619: contig of 24296 bp in length
*
* 144619 144719: gap of unknown length
*
* 144720 172326: contig of 27607 bp in length
*
* 172327 172426: gap of unknown length
*
* 172427 174046: contig of 1620 bp in length
*
* 174047 174146: gap of unknown length
*
* 174147 176137: contig of 1991 bp in length
*
* 176138 176237: gap of unknown length
*
* 176238 179816: contig of 3579 bp in length.
*
FEATURES
source
1. 179816
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-341G20"
misc_feature
1. 3838
/note="assembly_name:Contig10"
gap
3839. .3938
/estimated_length=unknown
misc_feature
3939. .9367
/note="assembly_name:Contig11"
gap
9368. .9467

```

misc\_feature /estimated\_length=unknown  
9468..15261  
/note="assembly\_name:Contig12"  
gap 15262..15361  
misc\_feature /estimated\_length=unknown  
15362..21615  
/note="assembly\_name:Contig13"  
gap 21616..21715  
misc\_feature /estimated\_length=unknown  
21716..28467  
/note="assembly\_name:Contig14"  
gap 28468..28567  
misc\_feature /estimated\_length=unknown  
28568..36770  
/note="assembly\_name:Contig15"  
gap 36771..36870  
misc\_feature /estimated\_length=unknown  
36871..48944  
/note="assembly\_name:Contig16"  
gap 48945..49044  
misc\_feature /estimated\_length=unknown  
49045..60729  
/note="assembly\_name:Contig17"  
gap 60730..60829  
misc\_feature /estimated\_length=unknown  
60830..74071  
/note="assembly\_name:Contig18"  
gap 74072..74171  
misc\_feature /estimated\_length=unknown  
74172..85808  
/note="assembly\_name:Contig19"  
gap 85809..85908  
misc\_feature /estimated\_length=unknown  
85909..101607  
/note="assembly\_name:Contig20"  
gap 101608..101707  
misc\_feature /estimated\_length=unknown  
101708..120221  
/note="assembly\_name:Contig21"  
clone\_end:SP6  
vector\_side:right"  
gap 120222..120321  
misc\_feature /estimated\_length=unknown  
120322..144619  
/note="assembly\_name:Contig22"  
gap 144620..144719  
misc\_feature /estimated\_length=unknown  
144720..172326  
/note="assembly\_name:Contig23"  
clone\_end:SP6  
vector\_side:right"  
gap 172327..172426  
misc\_feature /estimated\_length=unknown  
172427..174046  
/note="assembly\_name:Contig7"  
gap 174047..174146  
misc\_feature /estimated\_length=unknown  
174147..176137  
/note="assembly\_name:Contig8"  
gap 176138..176237  
misc\_feature /estimated\_length=unknown  
176238..179816  
/note="assembly\_name:Contig9"  
ORIGIN  
Query Match 26.3%; Score 526.6; DB 14; Length 179816;  
Best Local Similarity 78.7%; Pred. No. 5.7e-73;  
Matches 767; Conservative 0; Mismatches 169; Indels 39; Gaps 10;  
QY 1051 ATCACAAATTAAGTACTACA-CAAGTCTTTGGCTTCCAGTGCATATATAAGTTTGC 1109  
DB 149647 AACACAAAGAAAGCAAGTACATGAATGTTTGGTTTCCCAATGCAATATAAAGTTATGT 149706

QY 1110 TTATACTACGTAGTACTGTGTTAAGTGCATAG---TGTATGCTTAACAAAAACACAT 1166  
DB 149707 TTACACTATACGTAGACACAGAAAGTGCATAGATGTCTTAATATAACATATACAT 149766  
QY 1167 ACCTTAATTTTAAATGCTTTTATTAATAAAAAATGTAACAATC-ATTGAGCATTCAGT 1225  
DB 149767 ACTTAATTTTAAATAATCTTTATGCTTAATAAAAAATGCTGATATCAATCGAGCTTCGGT 149826  
QY 1226 GAGTGTATCTTTTGTGCTGTGGAAGGCTTTTCTTATGATGACTGAT-----1275  
DB 149827 GAGTGTACTCTTTTGTGCTGTGGAAGGCTTTGCTCGAGGTTGATGCTGACTGCT 149886  
QY 1276 -CGGGGCTCAGGCTGGAAGCTTAGGGTGGCTGTGGCAGTTTCTTAA-----ACAAG 1329  
DB 149887 CAGGCTGTGCTGCTGGAAGGTTGGGGTGGCTGTGGCAATTTCTTAAATTAAGACATCGG 149946  
QY 1330 TGAAGATTGCATATCATGTTGACTCTTCTTCATGAAAGATTTCTCTAGTGTGAT 1389  
DB 149947 TGAAGCTTTTGCATATGATTTGACTCTTCTTCATGAAAGATTTCTCTAGATGCGAT 150066  
QY 1390 GCTTTTGTATGATTTTATGCAAGTACACTCTTTGAAATTTGA-TCAATCCTCTC 1448  
DB 150007 GCTGTTTGTATGATTTTGTGCTGCAAGTACACTCTTTCATAAATGAGTCAATCCTCTC 150066  
QY 1449 AAACCTGCTGCTTTTAAACCTTAAGTTAATAATATTCGAAATTCATGTTGCTCAT 1508  
DB 150067 AAACCTGCTGCTGCTTTATCACTAAGTTAATAATATTCGAAATTCATGTTGCTCAT 150126  
QY 1509 TTCAACAATTTTACAGTGTCTTTCACAGAGATAGATTCATCTCATTTCTCGAGATGGA 1568  
DB 150127 TTCAACAATTTTACAGATCTTTCACAGAGATAGATTCATCTCATGAAACCACTT--- 150183  
QY 1569 ATCTTGTCTCATTCATAGAGAATAATCTCTCATCTGTTCAAGTTTATCATAGATTGCA 1628  
DB 150184 -TCTTTGCCCATTCATAGAGAATAAACAATCTCTCATCTGTTTAAAGTTTATCATAGATTGCA 150242  
QY 1629 GCAATACGTCAATGCTTCAGGCTCCTCATCTTTTATTCAGTCTCTGCTGTTTC 1688  
DB 150243 GCAATTCAGTCAATCTTTCAGG---CTTACTCTTAATTCAGTCTCTTGTGCTAATTTTC 150297  
QY 1689 TACCAATCTGTGTTCTTCTTCCATGTAAGTCTTGAACCTCTCAATTCATTCATGAG 1748  
DB 150298 CACCAATCTGTGACTTACTTCTTCCATGTAAGTCTTGAACCTCTCAATTCATTCATGAG 150357  
QY 1749 GGTGTGAATGCACTTCTTCAAAATTCCTGTTAATATTAATTAATTTGA---CCTCCATGA 1805  
DB 150358 GGTGTGAATGCACTTCTTCAAAATTCCTGTTAATATTAATTAATTTGA---CCTCCATGA 150417  
QY 1806 ATCAGGAATGTTCTTAATGCGACCTGGAATGCGTAATCCTTCCAAAGSTTTTCAATTT 1865  
DB 150418 ATCAGGAATGTTCTTAATGCGACCTGGAATGCGTAATCCTTCCAAAGSTTTTCAATTT 150477  
QY 1866 ACTTAGTCAGATTCATTCATCAGAGATTCACATTTCAATCCAGTTATAGCTTATAG 1925  
DB 150478 GCTTTGCCAGATTCAT-----CAGAGATCATCATTCATGCGCTAGACCTTATAGA 150532  
QY 1926 AATGATTTTCTCAATTAATAGGCTTGAAGTTGAATTAATCTCTTGATCATATTTTCTGC 1985  
DB 150533 AATGATTTTCTCAATTAATAGGCTTGAAGTTGAATTAATCTCTTGATCATATTTTCTGC 150592  
QY 1986 AAAATAGATGTTGTG 2000  
DB 150593 AGAATGATGTTGTG 150607  
RESULT 14  
AF307337  
LOCUS AF307337 210301 bp DNA linear PRI 23-APR-2001  
DEFINITION Homo sapiens ALPP, ALPP2, ALPI, XCB, CHRD, and CHRG genes,  
complete sequence.  
ACCESSION AF307337  
VERSION AF307337.1 GI:10732831  
KEYWORDS HTG.

QY	Db	Accession	Score	DB	Length	210301
QY	1051	ATCACAATAAAGGAGTCACACA-CAAGCTCTTTGGCTCCAGTGCATATATAAAGTTTGC	26.3%	Score 526.6;	DB 8;	Length 210301;
Db	208863	AACACAGAAAGCAAGTCACATGATATTTTGGTTCCCAATGACATATATAAAGTTATGT	Best Local Similarity 78.7%;	Pred. No. 5,4e-73;		
QY	1110	TTATACACACCTGTAGTCTGTGTAGTGTGCAATAG--TGTATGTCTAAAAAACAACAT	Matches 767; Conservative 0;	Mismatches 169;	Indels 39;	Gaps 10;
Db	208923	TTACACATATCTGTAGCCAGCAAGATGCAATATGATTTGTCTAATAAACAATATATCAT				
QY	1167	ACCTTAATTTTAAATATGCTTTATTTACTTAAAAATGCTAACATC-ATTGAGCATTTCACT				

[illegible]

Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hite, R.P., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Scherch, A., Schein, U.E., Jones, S.J., and Marra, M.A.

Mammalian Gene Collection Program Team  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932  
2 (bases 1 to 2641)

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
CONSRM  
JOURNAL  
TITLE  
REMARK  
COMMENT

NIH MGC Project  
Submitted (31-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadnan@systemsbio.org](mailto:amadnan@systemsbio.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAX Plate: 73 Row: j Column: 13  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22209000.

## FEATURES

source

1. 2641  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:43254 IMAGE:5264460"  
/tissue\_type="Brain, hippocampus"  
/clone\_lib="NIH MGC\_95"  
/lab\_host="DH10B"  
/note="Vector: pBluescriptR"  
780..2489  
/codon\_start=1  
/product="Unknown (protein for MGC:43254)"  
/protein\_id="AAH35143.1"  
/db\_xref="gi:38494046"

CDS

/translation="MTKSEBMSKREIRGLRQYVSQVNAKEKRLKVKASNP  
NTRNTRKNSLIADMEKLVVWIEDQTSRLPLSQSLQNFALTPNSWKEKGEA  
AEKFEASGFMREKERHFNIRKQGSAAADVAASAYEALAKIIDEQGYTKQ  
LFWDETAFWKMPKSRPTIAREKSPGFKSKDRLTLIGANAAQFKLPMILYH  
SENPRALKSYTKSTLPVLYKMSKAMTMTLFTAMTEFEKPTVEYCEKKIPKIL  
LIGANASHPRALMEIYEIRINYLPMANTSLLOPMDOGVSTPKSYLRNPFHALA  
AMDSPVDSGQSKTKTEPWKFTLIDAINIRDSWEVGLSTLTGWKKLITLIDDY  
BQKTSVEEVSADVIETAKELFEVREPVETLLOSHQDTLIDBELFLMDARKPFL  
MESTPGEDVAVIVEMTKDLEYINLVDAAGAFERIDSNFSSSTVGKMSNSTACY  
REIFHERKSQLMRKASPMYSFKLLPQPPSAATTLTQQPSTSRQDPPARVRLTE  
GSD"

ORIGIN

Query Match 26.2%; Score 523.4; DB 8; Length 2641;  
Best Local Similarity 79.0%; Pred. No. 8 6e-72;  
Matches 748; Conservative 0; Mismatches 161; Indels 38; Gaps 9;

QY	1078	TTTTGGCTTCCAGTCATATAAAGTTTGGCTTATATACACCTAGTCTGTAAAGT	1137
Db	2623	TTTTGGCTTCCAGTCATATAAAGTTTGGCTTATATACACCTAGTCTGTAAAGT	2564
QY	1138	GCATATG---TGTATGCTATAAAAAACATACCTTAATTTTAAATGCTTATAC	1194
Db	2563	GCATATGCAATTTGCTAATATAAACAATATACATACCTTAATTTTAAATGCTT	2504
QY	1195	AAAAATGCTAACAAATC-ATTGAGCATAGAGAGTGTATCTTTTGGTGGAGG	1253
Db	2503	AAAAATGCTAATTAATATCTAGCTTGGTGAATGTAATCTTTTGGTGGAGG	2444
QY	1254	TCTTTTCTTATGATGATGAT-----CGGGGGTCAAGTCTGAAGCTTAAAGG	1302
Db	2443	TCTTGGCTGAGGTTGATGATGTTGCTGACATGATGAGGCTGGTGGCTGAAGG	2384
QY	1303	GGCTGGCAGATTTCTTAAA-----ACAACAGTAAAGATTGCAATATCAGTTG	1357
Db	2383	GGCTGGCAGATTTCTTAAAATAAGACATCGGTGAAGCTTTTCGATGATGATCT	2324
QY	1358	CTTTCATGAAGAATTTCTCTAGTGTGATGATGATGATGATGATGATGATGAT	1417
Db	2323	CTTTCATGAAGAATTTCTCTGATGATGATGATGATGATGATGATGATGATGAT	2264
QY	1418	GAATCTCTTGAATAATGGA-TCATTCCTCTGAACCTCTGCTCTGCTTAAACA	1476
Db	2263	GAGCTCTTGAATAATGGAATGATGATGATGATGATGATGATGATGATGATGAT	2204
QY	1477	TTAATTAATTAATTCGATTCATATGCTGCTTCAATTAACAATTTTCAAGTGT	1536
Db	2203	TTAATTAATTAATTCGATTCATATGCTGCTTCAATTAACAATTTTCAAGTGT	2144
QY	1537	GGAGTATGATTCATCTCATCTATTCCTGAGATGATGATGATGATGATGATGAT	1596
Db	2143	GGAGTATGATTCATCTCATCTATTCCTGAGATGATGATGATGATGATGATGAT	2088
QY	1597	CTCATCTGTTCAAGTTTATTCATGATGATGATGATGATGATGATGATGATGAT	1656
Db	2087	CTCATCTGTTCAAGTTTATTCATGATGATGATGATGATGATGATGATGATGAT	2033
QY	1657	TTACCTTTAATTCAGTCTCTGCTGTTTCTACCAATCTGTGGTCTTCTCCATT	1716
Db	2032	TTACCTTTAATTCAGTCTCTGCTGTTTCTACCAATCTGTGGTCTTCTCCATT	1973
QY	1717	GAAGTCTGAACCTCTCCAGTCATCCATGAGAGGTTGAATGATCTTCCAAATTC	1776
Db	1972	GAAGTCTGAACCTCTCCAGTCATCCATGAGAGGTTGAATGATCTTCCAAATTC	1913
QY	1777	GTTAATTAATTAATTTGA---CTCCCATGATCATGATGATGATGATGATGATG	1833
Db	1912	GTTAATTAATTAATTTGAACCTCTCCCATGATCATGATGATGATGATGATGAT	1853
QY	1834	ATGATGAATCTTTCACAAAAGTTTCAATTTACTTATGTCAGATTCATTCAGAG	1893
Db	1852	ATGATGAATCTTTCACAAAAGTTTCAATTTACTTATGTCAGATTCATTCAGAG	1798
QY	1894	ATCCACTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1953
Db	1797	CATCATTCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1738
QY	1954	AAAGTGAATTAATCTCTGATGATGATGATGATGATGATGATGATGATGATG	2000
Db	1737	AAAGTGAATTAATCCCTTATGATGATGATGATGATGATGATGATGATGATG	1691

RESULT 16  
AL596257/c 60169 bp DNA linear PRI 18-MAY-2005

DEFINITION Human DNA sequence from clone RP11-251B9 on chromosome 1, complete sequence.  
ACCESSION AL596257 AC027619  
VERSION AL596257.5 GI:16973150  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 60169)  
Lad, H.  
AUTHORS Direct Submission  
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
JOURNAL  
COMMENT On Nov 16, 2001 this sequence version replaced gi:16304759.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP11-251B9 is from the library RPCR-11.1 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----  
Draft Sequence Produced by Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
http://genome.wustl.edu/gsc/index.shtml  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
FEATURES  
source  
1..60169  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-251B9"  
/clone\_1ib="RPCT-11.1"  
2000  
/note="Clone\_right\_end: RP4-614N24"  
58170  
/note="Clone\_left\_end: RP4-655C4"  
ORIGIN  
Query Match 26.1%; Score 522.2; DB 8; Length 60169;  
Best Local Similarity 75.8%; Pred. No. 4,1e-72;  
Matches 798; Conservative 0; Mismatches 198; Indels 57; Gaps 10;  
QY 976 GTACAGGCAATACCTTGAGATGCTGGGTTGGTTCATACACACCAATAAACAAT 1035  
|||||  
Db 48252 ATACAGCAATGCTCGAGATATTTGATGATTCAATCCACATCTAAGCAATAA 48193  
|||||  
QY 1036 ATGCAAGAAAGTGATATCAATTAAGTGATGACACACAATCTTTGGCTTCCAGTGA 1095  
|||||  
Db 48192 ATTG-----CAATTAAGCAATTCACATGATTTTGGTTTCCAGTGA 48148  
|||||

QY 1096 TATAAAGTTTGGCTATATACACTGAGCTGTTAGTGTGACATGATGTTATGCTA 1155  
|||||  
Db 48147 CATAAACTTATGTTTACATATACATGATGCTATATTAATCTGGCAATATGATATGCTA 48088  
|||||  
QY 1156 AAAAA--CACATACCTTAATTTTAAATGCTTATTAATCTAAATAATGCTAACATCAT 1212  
|||||  
Db 48087 AAACATGATATATCTTAATTTTAAATACTTTATTCCTGAAAAAAGCTATGCAATCAG 48028  
|||||  
QY 1213 TTGAGCATTCAGTGAAGTGTATCTTTTGGTGGTG-----GAAGTCTTTTCTTATTTGA 1267  
|||||  
Db 48027 CTGAGCTTCAGCGAGTGTATCTTTTGGCTGGAGTGAAGGGCTCTTGCTTGATGT 47968  
|||||  
QY 1268 TGACTGATCGGGGCTCA-----GGTGTGAACCTTAAAGGTGCTGGAGGCTT 1316  
|||||  
Db 47967 TGATGGCTGCTGACCTCACTCAGAGGATGAGTGTGGGAAAGTGTGTGACTGGCAATTT 47908  
|||||  
QY 1317 CTAA-----ACAACAGTGAAGATTGCAATATCAGTGACTCTTCTTCAAGAAAT 1371  
|||||  
Db 47907 CTTAAAGTAAACAAACAGAAAGTTTGGCTGATCCAGTGAACATTCCTTCAATGAAGAT 47848  
|||||  
QY 1372 TTCTCTGATGTGTGATGCTTTTGTATGACATTTATGACAGTGAACCTTTGAAA 1431  
|||||  
Db 47847 TTCTCTGATGATACATGCTGTGTGATGACATTTTACCACAGTGAACCTTTTCAAA 47788  
|||||  
QY 1432 ATTGA-TCATCTCTGAACCCCTGCTCTGTTTAAACAACCTAATTAATTAATTC 1490  
|||||  
Db 47787 ATTAACTGACGCTCTCAAAACCTGCTGCTTATCAACTGATGATTAATCAATTC 47728  
|||||  
QY 1491 TGAATCCATTTGTCATTTTCAACATTTTCAAGTGTCTTCCAGAGGATGATTCAT 1550  
|||||  
Db 47727 TAAATCCATGCTGCTCATTTTCAACAATGTTCAACATCTTCAACAGAGTATTCAT 47668  
|||||  
QY 1551 CTCATTTCTGATGATGAAATCTTTGCTCATTCATTAAGAAATTCCTCATGTTCAG 1610  
|||||  
Db 47667 CTTAAGAAACCACT--TCTTGTCTCATTAAGAAACCACTTCATCGTTAAG 47612  
|||||  
QY 1611 TTTTATCATGATGTCAGCAATATACATCATGCTTCAAGGCTCATCTTCACTTAATTC 1670  
|||||  
Db 47611 TTTTATCATGATGATGCAATATTCATGATCACTTCAGG-----CTCATCTTAATTC 47557  
|||||  
QY 1671 CAGTCTCTGCTGTTTCTTACACACATCTGTGCTCTTCTCATTTGAAGTCTGAACCT 1730  
|||||  
Db 47556 TGGTCTCTGCTTATTTTCCACACATCTGCAATCTTCTGTTCAACAATCTTGAACC 47497  
|||||  
QY 1731 CTCCAAGTCATCATGAGGTTGGAATGACCTTTCCAATTCCTGTTAATTTATAT 1790  
|||||  
Db 47496 CTCAAATATATCCGAGGATTTGAACCACTTCTTCAATTTCTGTCAAGTTGATAT 47437  
|||||  
QY 1791 TTTGAC--CTCCCATGATCATGATGTTCTTAATGACACTGGAATGTAATCTTT 1847  
|||||  
Db 47436 TTTGACCTTCTGCAATGATCATGATGATGTTCTTAATGGCATCTGAATGATGATCTTT 47377  
|||||  
QY 1848 CCAAAAGTTTCAATTAATTAATAGTCAGATCCATCCAGAGATCCACTTCAATG 1907  
|||||  
Db 47376 CCAAGAGTTTCAATTTCTTGGCCAGATCA-----CAAAGAGTCAATATTAAG 47322  
|||||  
QY 1908 CCAAGTATAGCCTTAATGAAATGATTTCTTCAATATTAAGGCTTGAAGTTGAATTAAT 1967  
|||||  
Db 47321 GAGCTCTAGCCTTCAAAATGATCTTTAATTAATTAAGCTGAAAGTCAAAATTAAT 47262  
|||||  
QY 1968 CCTTATCAATTTTCTGCAAAATGATGTTGG 2000  
|||||  
Db 47261 CCTGATCCATAGGCTGAGATTTGATGTTGG 47229  
|||||  
RESULT 17  
AC108017 51343 bp DNA linear PRI 13-MAY-2005  
LOCUS AC108017  
DEFINITION Homo sapiens BAC clone CTD-2005A22 from 4, complete sequence.  
ACCESSION AC108017  
VERSION AC108017.4 GI:19526131  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.

## REFERENCE

1 (bases 1 to 51343)  
Kozlowicz, A., Dignan, G. and Trani, L.  
The sequence of Homo sapiens BAC clone CTD-2005A22  
Unpublished (2001)

## JOURNAL

2 (bases 1 to 51343)  
Waterston, R.H.

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submision  
Submitted (24-JAN-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## REFERENCE

3 (bases 1 to 51343)  
Waterston, R.H.

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submision  
Submitted (23-FEB-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## JOURNAL

4 (bases 1 to 51343)  
Waterston, R.H.

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submision  
Submitted (16-MAR-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## REFERENCE

5 (bases 1 to 51343)  
Waterston, R.H.

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submision  
Submitted (27-MAR-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

## JOURNAL

6 (bases 1 to 51343)  
Wilson, R.K.

## AUTHORS

Wilson, R.K.

## TITLE

Direct Submision  
Submitted (13-MAY-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

## REFERENCE

On Mar 16, 2002 this sequence version replaced gi:18873881.

## COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@watson.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_MS2005A22  
-----

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
http://genome.wustl.edu

SOURCE INFORMATION:  
Clone CTD-2005A22 is from a release of the human BAC library CTD.  
The library contains cloned DNA from human sperm. See: Shizuya et  
al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al.,  
Genomics 34:213-8 (1996). The clone is available from Research  
Genetics, Inc. (http://www.resgen.com).  
VECTOR: pBeloBAC11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-719M18, 2000 bp overlap;  
the clone sequenced to the right is RP11-575B4, 2000 bp overlap.  
Actual start of this clone is at base position 120807 of  
RP11-719M18; actual end is at base position 7921 of RP11-575B4.

## FEATURES

Polymorphisms have been identified between AC108017 and AC019341.

## source

Location/Qualifiers  
1..51343

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/clone="CTD-2005A22"  
/clone\_1fb="CTD"  
/clone\_1lb="CTD"

## gene

7452..47527  
/gene="IGR7"  
/join(7452..7550,13377..13482,19119..19190,22023..22094,  
26181..26252,26342..26413,31182..31256,40871..40942,  
42689..42760,47456..47527)  
/gene="IGR7"

## CDS

join(<7452..7550,13377..13482,19119..19190,22023..22094,  
26181..26252,26342..26413,31182..31256,40871..40942,  
42689..42760,47456..47527)  
/product="unknown"  
/protein\_id="AAV41039.1"  
/db\_xref="GI:6395606"  
/translation="DNNMSLQFDKRYFASYKMTSQYPEAREPECLVSVPOCLQ  
GLELDCDETNLRAPSVSNVNTKMSLQWNLIRKLPDCKFYHDLQKYLQNNKITSI  
SLIVRGNLSLTLYLSHNRITPFRGVFEDLHRLMLIBDNHLSRISPTFGYLS  
LIIIVMNNVITRLPKDLCOHMPRIHWDLEGNRIHNLRIITFTSCNLTLYLVNRKN  
KINHNENTPAPLQDLDELGSNKIENPLIFDLKELSO"

## ORIGIN

Query Match 25.9%; Score 518.8; DB 8; Length 51343;  
Best Local Similarity 75.1%; Pred. No. 1.5e-71;

Matches 787; Conservative 0; Mismatches 207; Indels 54; Gaps 9;

```

QY 977 TACAGGCACTTCTGAGATGCTGGGGTGGTTCATACCAACAATAACAAATA 1036
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 26996 TACAGGCACTTCTGAGATGCTGGGGTGGTTCATACCAACAATAACAAATA 27055
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1037 TGCAGAAGTGATTCACATAAAGTGCACCAAGTCTTTGGCTGCCA---GCG 1093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27056 -----TCACAATAAAGCAAGTGCACCAAAATTTTGGTTCACGTGCA 27100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1094 CATATAAAGTTTCTTATCTACACGTGAGTCTGTTAAGTGCATAATGTTATGTC 1153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27101 CATATAAAGTTTCTTATCTACACGTGAGTCTGTTAAGTGCATAATGTTATGTC 27160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1154 TAAAAAACAATACCTTAATTTTAAATGCTTTATTAATAAATAAGCTAACAATCA 1213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27161 TAAAAAATGATACCTTAATTTTAAATGCTTTATGCTAATAAAGCTAACAATCA 27220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1214 TGAGCATTCAGTGAATGTAATCTTTTGGTGGTGAAGGCTTTTCTTATGATGAC 1273
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27221 TGAGTTTCAGGAGTCAATTAATCTTTTGGTGGTGAAGGCTTCATGCAATGTTGAC 27280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1274 ATCGGGGGTCA-----GGTGTGAAGCTTAGGGTGGTGGCAGTTTCTTAAA 1322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27281 CTGCTGACTTAACACAGTAGTGTTGTCGAAGCCGCGGGTGGTGGCAGTTTCTTAAA 27340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1333 -----ACAACAGTGAAGATTCGAATGCAATGCACTTCTCTTTCAGTAAGATTCTCT 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27341 ATTAAGATTAACAATGAGTTCATACATGATTTGATTTTAAATTCAGTAAGATTCTCT 27400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1378 CTAGTGTGATGCTTTTGTATGATTTTATGCAATAGCACTTTTGAATAATGGA 1437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db      27401 ATGGTAGCAATGCAATTTGTACATTTTACCAAGAGAAATTTCTTCAAAATGAA 27460
Qy      1438 -TCGAATCTCTCAACCCCTGCTGCTTAACACCTTAAGTAATATATTCGATC 1496
Db      27461 GTCAAATCTCTCAATCTCTCAATGCTTATCAATATATATATTCGATC 27520
Qy      1497 CATGTTGTCTCAATTTCAACATTTTCAAGTCTTCAACGAGAGATGATTCATCTCAT 1556
Db      27521 CTTTGTGTCTCAATTTCAACAGTCTTCAAGCATCTTACAGAGATGATTCATCTCA 27578
Qy      1557 TCCTGAGATGGAATCTTGTCTCATCATCAAGAGAAATTCCTCATCTGTCAAGTTTAT 1616
Db      27579 --AGAGAGCACTTCTTGTCTCATCATCAAGAGAAATTCCTCATCTGTCAAGTTTAT 27636
Qy      1617 CATGAGATTTGAGCAATACAGTCAATGCTTCAAGGCTCACTTCACTTTAATTCAGTTC 1676
Db      27637 CATGAGATTTGAGCAATACAGTCAATGCTTCAAGGCTCACTTCACTTTAATTCAGTTC 27691
Qy      1677 TCTTGTCTTCTCAACCATCTGTGCTTCTCTCATCAAGAGATCTTCAAGTCTTCAAG 1736
Db      27692 TCTTGTCTTCTCAACCATCTGTGCTTCTCTCATCAAGAGATCTTCAAGTCTTCAAG 27751
Qy      1737 GTCATTCATGAGGCTTGAATGCACTTCTTCAAAATTCCTGTTAATTTATTTTGAAC 1796
Db      27752 GTCACCCATGAGGCTTGAATGCACTTCTTCAAAATTCCTGTTAATTTATTTTGAAC 27811
Qy      1797 C----TCCCATGAATCAAGAAATGCTTAAATGAGCACTGGAATGGAATGCTTCCAAA 1852
Db      27812 CTTGTTTTCATGAATCAAGAAATGCTTAAATGAGCACTGGAATGGAATGCTTCCAAA 27871
Qy      1853 AGGTTTTCATTTACTTATGTCAGATCCATCCATCCAGAGATCCATCTTCAATGCCAGT 1912
Db      27872 AGGTTTTCATTTACTTATGTCAGATCCATCCATCCAGAGATCCATCTTCAATGCCAGT 27925
Qy      1913 TATAGCCTTATGGAATGATTTCTTCAATATATAGGCTTGAAGTTGAATTTACTCTCTG 1972
Db      27926 AGTACCTTTACAAAATGATCATCTTAAATATATAGCTTGAAGTTGAATTTACTCTCTG 27985
Qy      1973 ATCCATTTTCTGCAAAATATGATGTTGTG 2000
Db      27986 ATCCATGAGCTGTATGAAATGATGTCGTG 28013

RESULT 18
AC158447 135818 bp DNA linear HMG 12-APR-2005
LOCUS loxodonta africana clone VMRC15-156A5, WORKING DRAFT SEQUENCE, 4
DEFINITION ordered pieces.
ACCESSION AC158447.2 GI:62510111
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS loxodonta africana (African savanna elephant)
SOURCE loxodonta africana
ORGANISM Mammalia; Eutheria; Afrotheria; Proboscidea; Elephantidae;
loxodonta.
1 (bases 1 to 135818)
Antionellis,A., Ayele,K., Benjamin,B., Bikesley,R.W., Boakye,A.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H.,
Engle,J., Fukeenko,T., Gestole,M., Greene,A., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hunter,G.,
Hurle,B., Idol,J.R., Kwong,P., Latic,P., Larson,S., Lee-Jin,S.-Q.,
Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,
Mastaglio,C., Maskeri,B., McDowell,J., Mojidi,H.A., Mullikin,J.C.,
Oestreicher,J.S., Park,M., Portnoy,M.E., Prasad,A., Puri,O.,
Redix-Dugue,N., Schandler,K., Scheler,M.G., Sison,C.,
Stadripop,S., Stephen,B., Taye,A., Thomas,J.W., Thomas,P.J.,
Tsipouri,V., Ung,L., Vogt,J.L., Wehethy,K.D., Young,A. and
Green,E.D.
NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135818)

```

```

AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-2005) NIH Intramural Sequencing Center, 5625
Fishers Lane, Rockville, MD 20852, USA
REFERENCE 3 (bases 1 to 135818)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2005) NIH Intramural Sequencing Center, 5625
Fishers Lane, Rockville, MD 20852, USA
COMMENT On Apr 12, 2005 this sequence version replaced gi:61326250.
Genome Center

Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoehngri.nih.gov
----- Project Information -----
Center project name: jlu
Center clone name: 156A05

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is generally based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 135058 bases at least Q40
Consensus quality: 135283 bases at least Q30
Consensus quality: 135424 bases at least Q20
Insert size: 14200; agarose-fp
Insert size: 135518; sum-of-contigs
Quality coverage: 9.47x in Q20 bases; agarose-fp
Quality coverage: 9.92x in Q20 bases; sum-of-contigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 30739: contig of 30739 bp in length
30740 30839: gap of unknown length in length
30840 114090: contig of 83251 bp in length
114091 114190: gap of unknown length in length
114191 129767: contig of 15777 bp in length
129768 129867: gap of unknown length in length
129868 135818: contig of 5951 bp in length.
Location/Qualifiers
1..135818
/organism="loxodonta africana"
/mol_type="genomic DNA"
/db_xref="taxon:9785"
/clone="VMRC15-156A5"
/clone_1ib="VMRC15"
/note="BAC resource: http://bacpac.chori.org/"
1..30739
/note="assembly_fragment
clone_end:17
vector_side:left"
30740..30839
/estimated_length=unknown

```



	misc_feature	30840..114090	/note="assembly_fragment"
	misc_feature	70837..135818	/note="clone overlaps with GenBank Accession Number AC157635 clone VMRC15-551119 (center project name jlt)"
	gap	114091..114190	/estimated_length=unknown
	misc_feature	114191..129767	/note="assembly_fragment"
	gap	129768..129867	/estimated_length=unknown
	misc_feature	129868..135818	/note="assembly_fragment"
		clone_end:SP6	
		vector_side:right"	
ORIGIN			
Query Match	25.8%;	Score 516.8;	DB 14; Length 135818;
Best Local Similarity	76.2%;	Pred. No. 2.1e-71;	
Matches 802;	Conservative 0;	Mismatches 187;	Indels 63; Gaps 11;
Oy	970	TAGATAGTACAGGCATACCCTTGAGATGACTGGGGTTGGGTCATTACACACCAACAATAAT	1029
Dd	107981	TTGAATATPACAGSCATACCTTGAGAATATTGGCGAATTGGCTCCAGACCACCATATAA	108040
Oy	1030	ACAAATATGCCAAGAGTGATATCACATTAAGTAGTCACACAAAGCTTTTGGCTTCCC	1089
Dd	108041	GCAAAATATCGCA-----AGTACATGAGATTTTTTGGTTTACC	108077
Oy	1090	AGTGCAATATAAAGTTTTGCTTATACTACACGTAGTCTGTTAAGTGCATAATAGTTTA	1149
Dd	108078	AGTACATGTAAAAGTTATGTTTATGCTATTAATGATAGTCTATTBAGTGTCAATGACATTA	108137
Oy	1150	TGCTAAAAAAC-----ACATACCTTAATTTTAAATGCTTATTAATCTTAAAAATGCTA	1204
Dd	108138	TGCTTGA AAAACCAATGTACATACCTTAATTA AAAATACTTTATGCTAAAAATGCTA	108197
Oy	1205	ACAAATATTTAGGCATTCAGTGAATTTTATCTTTTGGCTGGTGAAGGCTTTTCTTAT	1264
Dd	108198	ACCATCATTCTGAGCCCTTGCAAGGTGATATCTTTTGGCTGGAGGGGCTTGCCCTTA	108257
Oy	1265	TGATGACTGAT-----CGGGGGTCAAGGCTGCTGAAGCTTAGGGTGGCTGTGGCAG	1313
Dd	108258	TGTTGATAGCTGCTGATCATGATCAGGGTGGTATCACTGAAGGTTGGGGTGGCTGTGGCAG	108317
Oy	1314	TTTCTTAAAAACA-AACAGTGAAGATGTCAAATACAGTGAAGCTTCTTTCATGAAGAAT	1371
Dd	108318	TTTCTTAAAAATAAGCAAAATCTGTTTGCCCCCATGTATGATGTTCTTTCATGAAGAAT	108377
Oy	1372	TTCTCTGATGCTGATGAGCTTTTGTATGCAATTTTATGCAACAGTGAAGCTTCTTGAAT	1431
Dd	108378	TTCTCTGATGATCGATGCTAATTTGATGATGCAATTTTCTCAAGCAAACTCTTTCAAA	108437
Oy	1432	AATGGA-TCAATCTCTCAAAACCTGCTGCTTGAACAACCTAAGTAAATATATATTC	1490
Dd	108438	AATGGAATCAATCTCTCAAGCCCTGCCATTTTATTTAA-CTAAGTTATATGAATATTC	108496
Oy	1491	TGAATCAATGTTGTCAATTTCAACAATTTTCAAGTGTCTTCAACAGAGATGATTCAT	1550
Dd	108497	TAAATCTTGTGTTCATTTCAACAACATTTCAAGCATTTTCAACAGAGATGATTCAT	108556
Oy	1551	CTCATTTCTGATGAGAACTTTTGCTATTCATAGAAGAATTCATCTGTCAAG	1610
Dd	108557	CTCAAGAAACCAAT---TCTTGTCTATTCATAGAAGAACTCCCATCTGTAAAG	108612
Oy	1611	TTTTATCATGAGATTCAGCAATACAGTATGTTTCAGGCTCACTTCACTTTTAATTC	1670
Dd	108613	TTTTATCATGAGATTCAGCAATACAGTATGTTTCAGGCTCACTTCACTTTTAATTC	108657
Oy	1671	CAGTCTCTGTGTTTCTTCAACAATCTGTGTTCTTCTCATTTGAAGTCTTGAACCT	1730
Dd	108668	TCTTCTCTGTGTTTCTTCAACAATCTGTGTTTCTTCTCATTTGAAGTCTTGAATCC	108727

QY	1731	TTCCAGTCATCCANTGAGGGTTGGAAATGACCTCTTCCAAATTCGCTTAATATTTATAT	1790
QY	108728	CTCAAGTCATCCAAAGGGCTGAATATCACTCTTCCAACTCCTGTCAGTTGAGGT	108787
QY	1791	TTTGA--CCTCCACATGATCATGATGTTCTTAAATGGCACCCTGAAATGATGATCCCTT	1847
Db	108788	TTTGACCTCCCTCCCATGATGATGATGATGTTCTTGATGACATCTAATATGTTGAATCCCTT	108847
QY	1848	CCAAAGGTTTTCATTTACTTAGTCCGATGCCATCCATCCAGAGATCCACTTTCAAATG	1907
Db	108848	TCAGAGGTTTTCATTTACTTTCCTGCCCCAGATCCAT-----CAGAGGAATCACCATCTGTG	108902
QY	1908	CCAGTTATAGCCTTATGCAATGATATTTCTTCATATATAAGCTTGAAAGTTGAAATTA	1967
Db	108903	GCAGCTATAGCCTTACAAATATGATATTTCTT---AATATAGACTTAAAGTTGAAATTA	108959
QY	1968	CCTGATTCATTTTCTGCAAAATGATGTTGT	1999
Db	108960	CTTTGACCCATGGCGCTGCAGATGATGATATGT	108991

	RESULT	19
Locus	ACI57635	
DEFINITION	Loxodonta africana clone VMCrl5-55119, WORKING DRAFT SEQUENCE, 6 ordered pieces.	
Accession	ACI57635	
VERSION	ACI57635.2	GI:62123027
KEYWORDS	HNG; HTGS_PHASE2; HTGS_DRAFT.	
SOURCE ORGANISM	<i>Loxodonta africana</i> <i>Loxodonta africana</i> ( <i>African savanna elephant</i> ) Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumalaya; Eutheria; Afrotheria; Proboscidea; Elephantidae; <i>Loxodontinae</i> .	
REFERENCE AUTHORS	1 (bases 1 to 141510) Antonecillis,A., Ayele,K., Benjamin,B., Blakesley,R.W., Boake,A., Boufield,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H., Engle,J.J., Fiksenko,T., Gestole,M., Greene,A., Guan,X., Gupta,J., Haghighi,P., Han,J.T., Hansen,N., Ho,S.-L., Hu,P., Hunter,G., Hurte,B., Idol,J.R., Kwong,P., LaRic,P., Larson,S., Lee-Lin,S.-Q., Legaepi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Matello,C.C., Maekeri,B., McDowell,U., Mojidi,H.A., Mullikin,J.C., Oestreicher,J.S., Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddi-Dugue,N., Schandler,K., Schneller,M.G., Sison,C., Stancitrop,S., Stephen,E., Taye,A., Thomas,J.W., Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L., Wehenky,K.D., Young,A. and Green,E.D. NISC Comparative Sequencing Initiative Unpublished 2 (bases 1 to 141510)	
JOURNAL REFERENCE TITLE	Green,E.D. Direct Submission Submitted (24-FEB-2005) NIH Intramural Sequencing Center, 5625 Fishers Lane, Rockville, MD 20852, USA 3 (bases 1 to 141510)	
JOURNAL REFERENCE TITLE	Green,E.D. Direct Submission Submitted (01-APR-2005) NIH Intramural Sequencing Center, 5625 Fishers Lane, Rockville, MD 20852, USA On Apr 1, 2005 this sequence version replaced gi:60223201.	
COMMENT		

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair

data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is generally based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

## ----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye, 100% of reads  
Assembly program: Phrap; version 0.99019  
Consensus quality: 140629 bases at least Q40  
Consensus quality: 140781 bases at least Q30  
Consensus quality: 140889 bases at least Q20  
Insert size: 148000; agarose-fp  
Insert size: 141010; sum-of-contigs  
Quality coverage: 9.15x in Q20 bases; agarose-fp  
Quality coverage: 9.61x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 42915: contig of 42915 bp in length  
\* 42916 43015: gap of unknown length  
\* 43016 58130: contig of 15115 bp in length  
\* 58131 58230: gap of unknown length  
\* 58231 88548: contig of 30318 bp in length  
\* 88549 88648: gap of unknown length  
\* 88649 106602: contig of 17954 bp in length  
\* 106603 114237: gap of unknown length  
\* 114238 114337: contig of 7535 bp in length  
\* 114338 141510: contig of 27173 bp in length.

## FEATURES

source  
1.141510  
/organism="Toxodonta africana"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9785"  
/clone\_lib="VMRC15-55119"  
/clone\_lib="VMRC15"  
/note="BAC resource: <http://bacpac.chori.org/>"  
1.6415  
/note="clone overlaps with GenBank Accession Number AC158447 clone VMRC15-156A5 (center project name j1u)"  
misc\_feature  
1.42915  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
42916..43015  
/estimated\_length=unknown  
43016..58130  
/note="assembly\_fragment"  
58131..58230  
/estimated\_length=unknown  
58231..88548  
/note="assembly\_fragment"  
88549..88648  
/estimated\_length=unknown  
88649..106602  
/note="assembly\_fragment"  
103152..141510  
/note="clone overlaps with GenBank Accession Number AC157633 clone VMRC15-499C6 (center project name j1a)"  
gap  
106603..106702  
/estimated\_length=unknown  
misc\_feature  
106703..114237

gap  
/note="assembly\_fragment"  
114238..114337  
/estimated\_length=unknown  
misc\_feature  
114338..141510  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:right"

## ORIGIN

Query Match 25.8%; Score 516.8; DB 14; Length 141510;  
Best Local Similarity 76.2%; Pred. No. 2e-71;  
Matches 802; Conservative 0; Mismatches 187; Indels 63; Gaps 11;

970 TAGATAGTACAGGCTATCTTGAGATACCTGCTGTTGTTCCATACCAACAATAT 1029  
Db TTGAATATACAGGCTATCTTGAGATATTCGGATTCGGCTCCAGACACACATATA 37203  
1030 ACAATATGCAAGAGTGCATATCAATTAAGTACACACAGCTTTGGCTCC 1089  
Db GCAAATATCGCA-----AGTACATGAGATTTTGGTTACC 37240  
1090 AGTGCATATTAAGTTTGTCTTATCTACACTGATGCTGTTAAGTGCATATGTTA 1149  
Db AGTACATGATTAAGTTATGTTATATGTTATGTTATTAAGTGCATATGTTA 37300  
1150 TGTCTTAAAAAC-----ACATCTTAATTTTAAATGCTTTATTAATAAATGCTA 1204  
Db TGTCTTAAAAACCAATGATACATCTTATTAATAAATGCTTTATGCTTAAAAATGCTA 37360  
1205 ACAATATTTGAGCTTACAGTACAGTGTATATCTTTGCTGTTGAGAGTCTTTCTAT 1264  
Db ACCATCATCTGAGCTTGTGCAAGTGTATCTTTTCTGTTGAGAGGCTTGTGCTTGA 37420  
1265 TGTGATGATGAT-----CGGGGTCAAGTGTCTGAAAGCTTGTGAGGCTGTGCGAG 1313  
Db TGTGATGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 37480  
1314 TTTCTTAAAAAC--ACAGTAAAGATGTGCAATATGATGATGATGATGATGATGATG 1371  
Db TTTCTTAAAAATGAAACAATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 37540  
1372 TTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1431  
Db TTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 37600  
1374 ATTGGA-TCATCTCTGCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1490  
Db ATTGGAATCATCTCTCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 37659  
1491 TGAATCATGTTGTTGATTTCAAAATTTTCAAGTGTCTTCAAGAGTATGATTCAT 1550  
Db TGAATCATGTTGTTGATTTCAAAATTTTCAAGTGTCTTCAAGAGTATGATTCAT 37719  
1551 CTCATTTCTGATGATGATTTTGTCTATTCATTAAGAAATTCCTATCTGTTCAAG 1610  
Db CTCATTTCTGATGATTTTGTCTATTCATTAAGAAATTCCTATCTGTTCAAG 37775  
1611 TTTTATCATGATTTGACGACATPATCAGTATCTTCAAGGCTCACTTCACTTTTATTC 1670  
Db TTTTATCATGATTTGACGACATPATCAGTATCTTCAAGGCTCACTTCACTTTTATTC 37830  
1671 CAGTCTCTGCTGTTTCTACCAATCTGTTGCTTCTTCTTCAATGAGTCTTGAACCT 1730  
Db TTTTATCATGATTTGACGACATPATCAGTATCTTCAAGGCTCACTTCACTTTTATTC 37890  
1731 CTCGAATCATCATGAGGTTGAAATGCACTTTCTTCAAAATTCCTGTTAATTTATAT 1790  
Db CTCGAATCATCATGAGGTTGAAATGCACTTTCTTCAAAATTCCTGTTAATTTATAT 37950  
1791 TTGGA---CTTCCATGATATGATGATGATGATGATGATGATGATGATGATGATGATG 1847  
Db TTGGAATCATCATGAGGTTGAAATGCACTTTCTTCAAAATTCCTGTTAATTTATAT 38010  
37951 TTTGACCTCTCCCATGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 38010

QY 1848 CCAAAAGTTTCAATTACTTATGTCAGATCCATCCAGAGATCCATTGCAAG 1907  
 DB 38011 TAGAAGGTTTCAATTACTTATGTCAGATCCAT-----CAGAGAAATCCATCTG 38065  
 QY 1908 CCAATTATGCTTATGATGATGATTTCTTCAATATTAAGGCTTGAAGTTGAATTACT 1967  
 DB 38066 GAGGATATGCTTATGATGATGATTTCTTCAATATTAAGGCTTGAAGTTGAATTACT 38122  
 QY 1968 CCTGATCCATTTTCTGCAAAATAGATGTTGT 1999  
 DB 38123 CCTGATCCATTTGCTGCAAGATGATTAAGT 38154  
 RESULT 20  
 AX833395  
 LOCUS AX833395 3895 bp DNA linear PAT 15-DEC-2003  
 DEFINITION Sequence 519 from Patent EP1347046.  
 ACCESSION AX833395  
 VERSION AX833395.1 GI:39919530  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE  
 AUTHORS 1  
 TITLE Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
 Yamamoto, J., I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
 Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and  
 Masuko, Y.  
 JOURNAL Full-length cDNA sequences  
 Patent: EP 1347046-A 519 24-SEP-2003;  
 Research Association for Biotechnology (JP)  
 FEATURES  
 source  
 1. .3895  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Query Match 25.8%; Score 516; DB 6; Length 3895;  
 Best Local Similarity 77.5%; Pred. No. 1e-70;  
 Matches 741; Conservative 0; Mismatches 175; Indels 40; Gaps 8;  
 QY 1071 ACAAAGCTTTGGCTCCAGAGCATATTAAGTTTGGTTATCTACCTGATCTGT 1130  
 DB 1362 ATTAATTTTGGTCCAGATGATTAAGTTTAAATTAATTAATTAATTAATTAAT 1421  
 QY 1131 TAAGTGTGCAATAGTGTATGTTCTTAAAAAA-----CACATACCTTAATTTTAAATGC 1184  
 DB 1422 TAAAGGTGCAATAGTGTATGTTCTTAAAAAAAACATGTCATGCTTAATTTAAATATGC 1481  
 QY 1185 TTTATTTACTTAAATAATGCTTAACATCATTTGAGCATTCAGTATGTTATCTTTTGGCT 1244  
 DB 1482 TTTATTTACTTAAATAATGCTTAACATCATTCGAACTTCAAGTATGTTATTTTGGCT 1541  
 QY 1245 GGTGGAAGTCTTTTCTTATTTGATGATCTAT-----CGGGGCTCAGGCTGTA 1293  
 DB 1542 AGTGAAGGCTCTTGGCTGATGATGATGCTGCTGATGATGATGATGATGATGATGAT 1601  
 QY 1294 GCTTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1348  
 DB 1602 GGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1661  
 QY 1349 TGACTCTTCTTCAATGAAGAATTTCTCTAGTGTGATGCTTTTGTATGATGATTTTA 1408  
 DB 1662 TGACTCTTCTTCAATGAAGAATTTCTCTAGTGTGATGCTTTTGTATGATGATTTTA 1721  
 QY 1409 TGCACAGTGAAGTCTTTTGAATAATGGA-TCAATCTCTCAAACTCTGCTCTGCTTTAA 1467  
 DB 1722 CCCACAGTGAAGTCTTTTGAATAATGGAAGTCAAGTCTCTTAAACCTCTGCACTCTTTTA 1781  
 QY 1468 CAACCTAAGTGAATTAATTAATTTCTGAATCCATTTGTTGATTTCAACATTTTCAAGAG 1527

DB 1782 TTAAGTGAAGTCAAGCAATATTTCTAATCTTTGTTGTCATTTCAACAATGTTCAACA 1841  
 QY 1528 TCTTCACACGAGATGATATTCATCTTCTTCTGAGATGGAATCTTTGTCATTCATAG 1587  
 DB 1842 TCTTCACACGAGATGATATTCATCTTCTGAGAAACCACTT-----TCTTTGTCATTCATAG 1897  
 QY 1588 AAGAAATTCCTCATCTGTTCAAGTTTATCATGATGATGATGATGATGATGATGATGATGAT 1647  
 DB 1898 AAGCAATCCCTCATCTGTTCAAGTTTATCATGATGATGATGATGATGATGATGATGATGAT 1957  
 QY 1648 AGGCTCATCTTCACTTTTAAATTCAGATCTCTGCTGCTTTCACACATCTGCTGCTTCT 1707  
 DB 1958 AGG-----CTGCAATTTTAAATTTCTAGTCTCTGCTGCTTTCACACATCTGCAATCT 2012  
 QY 1708 TCTTCATTTGAAGTCTTGAAGTCTTCAAGTATCATGATGATGATGATGATGATGATGATGAT 1767  
 DB 2013 TCTTCATTTGAAGTCTTGAAGTCTTCAAGTATCATGATGATGATGATGATGATGATGATGAT 2072  
 QY 1768 CAAATTCCTGTTAATTAATTAATTTTGAAGTCTTCAAGTATCATGATGATGATGATGATGATGAT 1824  
 DB 2073 CAAATTCCTGTTAATTAATTAATTTTGAAGTCTTCAAGTATCATGATGATGATGATGATGATGAT 2132  
 QY 1825 GCACCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1884  
 DB 2133 ACATCTAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2190  
 QY 1885 ATCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1944  
 DB 2191 ---CAGAGAAATCACGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2247  
 QY 1945 AAGCTTGAAGTGAATTAATTAATTTCTGATGATGATGATGATGATGATGATGATGATGAT 2000  
 DB 2248 ATGCTTGAAGTGAATTAATTAATTTCTGATGATGATGATGATGATGATGATGATGATGAT 2303  
 RESULT 21  
 AK095077  
 LOCUS AK095077 3895 bp mRNA linear PRI 30-JAN-2004  
 DEFINITION Homo sapiens cDNA FLJ37758 f18, clone BRHIP2023869.  
 ACCESSION AK095077  
 VERSION AK095077.1 GI:21754267  
 KEYWORDS oligo capping; f18 (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE  
 AUTHORS 1  
 TITLE Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
 Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
 Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
 Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
 Nagahari, K., Murakami, K., Yasuda, T., Iwano, T., Watanabe, M.,  
 Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
 Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
 Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
 Tanikawa, M., Yamazaki, M., Nishimura, K., Ishibashi, T., Yamashita, H.,  
 Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
 Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
 Yoisida, M., Houchi, T., Kusano, J., Kanohori, K., Takahashi, F.,  
 Hara, H., Tanabe, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,  
 Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, H., Oshima, A.,  
 Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,  
 Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,  
 Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
 Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
 Kunigai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
 Tashiro, H., Tanigami, A., Fujiiwara, T., Ono, T., Yamada, K., Fujii, Y.,  
 Ozaki, K., Hiro, M., Ohnori, Y., Kawabata, A., Hiki, T., Kobatake, N.,  
 Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
 Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Takagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
 Nat. Genet. 36 (1), 40-45 (2004)  
 14702039  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS  
 2  
 Niimiya, K., Magatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuma, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 3 (bases 1 to 3895)  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 Isogai, T. and Yamamoto, J.  
 Direct Submission  
 Submitted (04-JUN-2002) Takao Isogai, FUD Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomicehri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
 Location/Qualifiers  
 source  
 1..3895  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="BRHIP2023869"  
 /tissue\_type="hippocampus"  
 /clone\_id="BRHIP2"  
 /note="cloning vector: pME18SFL3"  
 ORIGIN  
 Query Match 25.8%; Score 516; DB 8; Length 3895;  
 Best Local Similarity 77.5%; Pred. No. 1e-70;  
 Matches 741; Conservative 0; Mismatches 175; Indels 40; Gaps 8;  
 QY 1071 ACAAGCTTTGGCTCCAGTGCATATTAAGTTTGGCTTATTCACAGCTGAGCTGT 1130  
 DB 1362 ATTAATTTTGGTTCCGACATGTAAGTTAAATTAATTAATTAATTAATTAATTA 1421  
 QY 1131 TAAAGTGCATATGCTTATGCTTAAAAA-----CATATACCTTAATTTAAAAATGC 1184  
 DB 1422 TAAAGTGCATATGCTTATGCTTAAAAAACAATGCTATGCTTAAATTAATTAATTA 1481  
 QY 1185 TTTTATTAATAAATGCTTAAACATATTTGACATTCAGTGAGTTGTAATCTTTTGCT 1244  
 DB 1482 TTTTATTAATAAATGCTTAAACATATTCGAACTTCAGTGAGTATATATTTTGGCT 1541  
 QY 1245 GGTGGAAGTCTTTTCTTATGATGCTAT-----CGGGGGTCAAGTGTCTGAA 1293  
 DB 1542 AGTGGAGGCTCTGCTGATGCTGATGCTGCTGATGCTGATGCTGATGCTGAA 1601  
 QY 1294 GCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1348  
 DB 1602 GGTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1661  
 QY 1349 TGAATCTTCTTCAATGAAGATTTCTCTAGTGTGAGTCTTTTGTATGATGATTTTA 1408  
 DB 1662 TGAATCTTCTTCAATGAAGATTTCTCTAGTGTGAGTCTTTTGTATGATGATTTTA 1721

QY 1409 TGCACAGTGAATCTTTTGAATAATGGA-TCATCTCTCAAACTGCTGCTTTAA 1467  
 DB 1722 CCCACAGTGAATTTATTTGAAATGAAAGCAGCTCTTAAACCTGCTGCTTTTA 1781  
 QY 1468 CAACCTAAGTTAATTAATATTTGCAATCCATGTTGTCATTTTCAACAAATTTCAAGT 1527  
 DB 1782 TTAATTAAGTTACAGCAATATTTTAAATCTTTGTTGTCATTTTCAACATGTCACACA 1841  
 QY 1528 TCTTCACAGAGATGATTCATCTCATTTCTGATGATGATGATGATGATGATGATGATGAT 1587  
 DB 1842 TCTTCACAGAGATGATTCATCTCATTTCTGATGATGATGATGATGATGATGATGATGAT 1897  
 QY 1588 AAGAAATTCCTATCTGTTCAAGTTTATTCATGATGATGATGATGATGATGATGATGATGAT 1647  
 DB 1898 AAGCAACCTCATCTGTTCAAGTTTATTCATGATGATGATGATGATGATGATGATGATGAT 1957  
 QY 1648 AGGCTTCACTTCATTTTAAATTCAGTTCCTGCTGTTTCTACACATCTGCTGCTTCT 1707  
 DB 1958 AGG-----CTCATTTTAAATTTAGTTCCTGCTGATTTTACCAATCTGCAATGAT 2012  
 QY 1708 TCTTCATGAAAGTCTGAACTCTCAAGTCAATGATGATGATGATGATGATGATGATGATGAT 1767  
 DB 2013 TCTTCATGAAAGTCTGAACTCTCAAGTCAATGATGATGATGATGATGATGATGATGATGAT 2072  
 QY 1768 CAATTCCTGTTAATTAATTAATTTTGAATCTCC--CATGAATCAATGATGATGATGATGAT 1824  
 DB 2073 CAATTCCTGTTAATTAATTAATTTTGAATCTCTCTCATGATGATGATGATGATGATGATGAT 2132  
 QY 1825 GCACTGGAATGATGATATCCCTTCCAAAGTTTCAATTTAATTTAATTTAATTTAATTTAAT 1884  
 DB 2133 ACATCTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2190  
 QY 1885 ATCCAGAGATCACTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1944  
 DB 2191 ---CAGAGATCACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2247  
 QY 1945 AAGGCTTGAAGTGAATTAATCTCTTGAATTCATTTTTCGAAATGATGATGATGATGAT 2000  
 DB 2248 ATGCTTGAAGTGAATTAATGATCTCTGATGATGATGATGATGATGATGATGATGATGAT 2303  
 RESULT 22  
 AC004736/c 115958 bp DNA linear PRI 03-JUN-1998  
 LOCUS Human Chromosome 11p14.3 PAC clone pJ1082L12 containing KNCN1 and  
 DEFINITION MyD, complete sequence.  
 AC004736 AC004736.1 GI:3282183  
 VERSION  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 115958)  
 Evans, G.A., Achanasios, M., Aguayo, P., Armstrong, D., Basil, M.,  
 Buettner, J., Bumelster, R., Card, P., desallboat, C., Dunn, J.,  
 English, C., Ehrbridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,  
 Grant, O., Hahner, L., Harris, J., Lewis, E., Loo, H., Loo, K.N.,  
 Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S.,  
 Schegeman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.  
 HTG Submission  
 Unpublished  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 115958)  
 Evans, G.A., Achanasios, M., Aguayo, P., Armstrong, D., Basil, M.,  
 Buettner, J., Bumelster, R., Card, P., desallboat, C., Dunn, J.,  
 English, C., Ehrbridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,  
 Grant, O., Hahner, L., Harris, J., Lewis, E., Loo, H., Loo, K.N.,  
 McFarland, J., Newton, J., Osborne-Lawrence, S.,  
 Schegeman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.  
 Direct Submission  
 Submitted (23-MAY-1998) Genome Science & Technology Center,

University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA

REFERENCE  
AUTHORS  
Evans,G.A., Achanaslou,M., Aguayo,P., Armstrong,D., Basil,M., Buetner,J., Bumeister,R., Card,P., desaliboat,F., Dunn,J., English,C., Enridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.

TITLE  
JOURNAL  
Submitted (01-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA

REFERENCE  
AUTHORS  
Evans,G.A., Achanaslou,M., Aguayo,P., Armstrong,D., Basil,M., Buetner,J., Bumeister,R., Card,P., desaliboat,F., Dunn,J., English,C., Enridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.

TITLE  
JOURNAL  
Submitted (03-JUL-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA

COMMENT  
NOTE: Extent of PDJ1082L12 overlap with mapped clones: SP6 PAC cloning end: 1. 8775, PDJ59m18; T7 PAC cloning end: 45252. 115958, 6-109h6. Further information regarding the map of this region or annotation of PDJ1082L12 can be found at <http://geotec.swmed.edu/chromoso2.htm>.  
CHROMOSOMAL LOCUS: This PAC clone comes from the Uhler syndrome region (USH1C) mapped between STS markers D11S1310 and 115A14.  
MARKER CONFIRMATION: STS/ESTs sequence confirmed; D11S4130, KNCN1, and MYO1L.  
MAPPED CLONE OVERLAP: PDJ59m18 and 6-109h6.  
IMPORTANT: This submission contains the entire insert of clone PDJ1082L12. PDJ1082L12 comes from a PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.39. In addition, this sequence has been finished such that 99.9% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.

FEATURES  
SOURCE  
1. 115958  
/organism="Homo sapiens"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:9606"

repeat\_region  
/rpt\_family="MIR"  
4405..4517

repeat\_region  
/rpt\_family="Alu"  
complement(19171..19291)

repeat\_region  
/rpt\_family="Alu"  
complement(21561..21985)

repeat\_region  
/rpt\_family="Alu"  
23678..23767

repeat\_region  
/rpt\_family="MIR"  
28877..31250

repeat\_region  
/rpt\_family="Tigger1"  
complement(31777..31876)

repeat\_region  
/rpt\_family="MIR"  
complement(36023..36231)

repeat\_region  
/rpt\_family="MIR"  
complement(39712..39766)

repeat\_region  
/rpt\_family="MIR"  
complement(41971..42073)

repeat\_region  
/rpt\_family="MIR"  
45642..45752

repeat\_region  
/rpt\_family="MIR"  
57204..57337

repeat\_region  
/rpt\_family="MIR"  
complement(58027..58249)

repeat\_region  
/rpt\_family="MER20"  
59316..59729

repeat\_region  
/rpt\_family="L1"  
complement(60013..60386)

repeat\_region  
/rpt\_family="THE1"  
60398..60596

repeat\_region  
/rpt\_family="L1"  
60940..61228

repeat\_region  
/rpt\_family="Alu"  
complement(63635..63900)

repeat\_region  
/rpt\_family="Alu"  
complement(75510..75774)

repeat\_region  
/rpt\_family="Alu"  
complement(82106..82378)

repeat\_region  
/rpt\_family="Alu"  
complement(84150..84211)

repeat\_region  
/rpt\_family="MLT1"  
84727..85172

repeat\_region  
/rpt\_family="MLT1"  
85852..86134

repeat\_region  
/rpt\_family="Alu"  
86387..86565

repeat\_region  
/rpt\_family="Alu"  
complement(88437..88514)

repeat\_region  
/rpt\_family="MIR"  
complement(90686..90752)

repeat\_region  
/rpt\_family="MIR"  
91042..91163

repeat\_region  
/rpt\_family="MIR"  
94274..94419

repeat\_region  
/rpt\_family="MIR"  
complement(95470..95617)

repeat\_region  
/rpt\_family="MLT1"  
complement(97484..97917)

repeat\_region  
/rpt\_family="L1"  
105586..105657

repeat\_region  
/rpt\_family="MIR"  
105673..105884

repeat\_region  
/rpt\_family="MIR"  
complement(106369..106641)

repeat\_region  
/rpt\_family="Alu"  
complement(107302..107596)

repeat\_region  
/rpt\_family="Alu"  
109188..109458

repeat\_region  
/rpt\_family="Alu"  
complement(110473..110773)

repeat\_region  
/rpt\_family="Alu"  
complement(111227..111337)

repeat\_region  
111412..111792

repeat\_region  
/rpt\_family="THE1"  
complement(112629..112992)

repeat\_region  
/rpt\_family="L1"

ORIGIN  
Query Match 25.8%; Score 515.2; DB 8; Length 115958;  
Best Local Similarity 75.6%; Pred.No.3.9e-71;  
Matches 789; Conservative 0; Mismatches 203; Indels 52; Gaps 10;  
977 TACAGCATCTCGAGATCTGAGGTTGGTTCCATACACACAATATATCAATTA 1036  
|||||  
31252 TACAGCATCTCGGATATTTGTGGCTTGTTCCAGACACAGCAATA----- 31202  
1037 TGCAAGAAGTGATATCAATTAAGTAGTACACACAAGTCTTTGGCTTCCAGTGAT 1096  
31201 -----AAGTAATATCAATTAAGCAAGTACATAA-----TTTTGGTTTCTCAGAGCAT 31150  
1097 ATAAAGTTTGCTTATCTACACTGAGTGTGTTAGTGCAATGTATATGCTAA 1156  
|||||  
31149 ATAAAGTTTATTTATCTACTGATATTAATGCAATGCAATTATGCTAA 31090  
1157 AAAACACATACCTTAATTTTAAATGCTTATTTACTAAAAAATGCTAACATGATTGA 1216  
|||||

```

Db      31089 AATGTACATATCTTAAATTTTAAATAGTTATGCTTCAAAATGCTAACAATATATdA 31030
Qy      1217 GCATTACGTAGTGTATCTTTTGTGCTGAGGAGCTTTCTTATTTGATGACTAT- 1275
Db      31029 GCCTTCAGTGTGATATCTTTTGTGCTGAGGAGGCTTCTCAGTGTGATGAGCTG 30970
Qy      1276 -----CGGGGGTCAGGCTGAGGAGCTTAAGGGTGGCTGTGGCAGTTCTTAAA--- 1322
Db      30969 CCGATTGATCAGCGGTGATGCTGAGGAGGTTGGGGGTGACATAGCAATTTGTAAAGTA 30910
Qy      1323 --ACACAGTAAAGTTGCAATATGATGACTCTCTCTTTCATGAAAGATTTCTCTTA 1380
Db      30909 AGACAACAAGAGTTGCTGATCAATTTGATCTCCCTTTTCAAAAAGATTTCTCTTA 30850
Qy      1381 GTGTGTATGCTTTTGTATGATTTTGTAGATTTTGTAGACAGTAACTTTTGAATTTGA--TC 1439
Db      30849 GCATGTATGCTTTGTTGACAGATTTCCATCCACATGAACTTTCTTCAAAAATGAAATGTC 30790
Qy      1440 AATCTCTCAAAACCTGCTGCTTTTAACAACCTAAGTTAATATAATTTGAAATGCAT 1499
Db      30789 AATCTCTCAAAACCTGCTGCTTTTATCAGCTGATTTATGAAATATTTAAATCTCT 30730
Qy      1500 TGTGTCTATTTCAACAATTTTCAAGTGTCTTACCAGAGATGATTTCAATCTCATTTCC 1559
Db      30729 TGTGTCTATTTCAACAATGTTATATGATCTTACACAGAAATATGATTCATCTCAAGAA 30670
Qy      1560 TGAGATGGAATCTTGTCTCATCATAGAGAAATTCCTCATCTGTTCAAGTTATCAT 1619
Db      30669 CCACTT---TCTTGTCTTATCATAGAGAACTCTCATCTGTTCAAGTTATCAT 30614
Qy      1620 GAGATGAGAGCAATACAGTCAATGCTTTCAGGCTTCACTTCAATTTTCAAGTTCT 1679
Db      30613 GAGATGAGAGCAATACAGTCAATGCTTTCAGGCTTTCAGGCTTTCAGGCTTCT 30559
Qy      1680 TGTCTTCTTCAACAATCTGCTGCTTCTTCTTCAATGAACTTGAACCTTCCAAGTC 1739
Db      30558 TGTCTTCTTCAACAATCTGCTGCTTCTTCTTCAATGAACTTGAACCTTCCAAGTC 30499
Qy      1740 ATCCATGAGAGGTTGAATGAGCTTCTTCAAAATTCCTGTTAATTTATTTTGA---C 1796
Db      30498 ATTCAGAGAAATGGAATGAACTCTTCTTGAACCTGTTAATGTTGCTGTTTGAAGCTC 30439
Qy      1797 CTCCTGATGATCAATGATTTCTTAAATGAGCAGCTTGAATGCTTTCAAAAGGT 1856
Db      30438 CTCCTGATGATCAATGATTTCTTAAATGAGCAGCTTGAATGCTTTCAAAAGGT 30380
Qy      1857 TTTCAATTTACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1916
Db      30379 TTTCAATTTACTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 30325
Qy      1917 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1976
Db      30324 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 30265
Qy      1977 ATTTCTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2000
Db      30264 ATGGGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 30241

```

```

RESULT 23
AC124056 115971 bp DNA 1linear PRI 21-DEC-2002
DEFINITION Homo sapiens chromosome 11, clone RP5-1082L12, complete sequence.
ACCESSION AC124056
VERSION AC124056.8 GI:27356724
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 115971)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.

```

# TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens chromosome 11, clone RP5-1082L12  
Unpublished  
2 (bases 1 to 115971)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
Anderson,S., Barna,N., Baetien,V., Bloom,T., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Gird,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamet,A., Karatas,A., Kelle,C., LaRocque,K.,  
Lamazares,R., Landers,T., Lenockky,J., Levine,R., Lindblad-Toh,K.,  
Lin,G., Maclean,C., Macdonald,P., Major,J., Margulis,N.,  
Matthews,C., McCarthy,M., McKernan,K., Meldrum,J.,  
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,  
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Pollara,V., Raymond,C., Retta,K., Rieback,M., Riley,R., Rise,C.,  
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,  
Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testfaye,S.,  
Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

# TITLE JOURNAL REFERENCE AUTHORS

Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 115971)  
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gird,S., Graham,L., Grand-Pierre,N., Hafiz,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamet,A., Karatas,A., Kelle,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Lin,G., Maclean,C., Macdonald,P., Major,J.,  
Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,  
Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,K.,  
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Talamas,J., Testfaye,S., Theodore,J., Topham,K.,  
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 21, 2002 this sequence version replaced gi:25102226.

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L27400  
Center clone name: 1082\_L\_12

## FEATURES

## source

```
Location/Qualifiers
1..115971
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP5-1082L12"
/clone_lib="RPCL Human PAC library 5"
repeat_region
99..1108
/rpt_family="L2"
repeat_region
1206..1473
/rpt_family="L2"
repeat_region
complement(1602..1704)
/rpt_family="L1ME"
repeat_region
complement(1845..2046)
/rpt_family="MIR3"
repeat_region
complement(2352..2469)
/rpt_family="MIR"
repeat_region
2612..4114
/rpt_family="LMC1"
repeat_region
4083..4170
/rpt_family="L1MEB3A"
repeat_region
complement(4171..4551)
/rpt_family="MSTA"
repeat_region
4552..4920
/rpt_family="L1MEB3A"
repeat_region
4925..4986
/rpt_family="L1MEB3A"
repeat_region
4970..5179
/rpt_family="L1MD3"
repeat_region
5180..5490
/rpt_family="AluSx"
repeat_region
5491..5609
/rpt_family="L1MD3"
repeat_region
5609..5845
/rpt_family="L1MEB3A"
repeat_region
complement(6442..6757)
/rpt_family="AluSx"
repeat_region
8356..8671
/rpt_family="AluSx"
repeat_region
8717..8725
/rpt_family="AluSx"
unsure
8735..8770
/note="<30 qual SNGL region"
unsure
8786..8790
/note="<30 qual SNGL region"
unsure
9317..9609
/rpt_family="AluSc"
repeat_region
complement(10070..10405)
/rpt_family="MIR"
repeat_region
complement(11554..11859)
/rpt_family="L1R16C"
repeat_region
12223..12374
/rpt_family="MIR"
repeat_region
13250..13353
/rpt_family="MIR3"
repeat_region
complement(14122..14358)
/rpt_family="MIR"
repeat_region
14653..14758
/rpt_family="MIR"
```

```
repeat_region 16324..17416
/rpt_family="L2"
repeat_region complement(17963..18020)
/rpt_family="MLT1B2"
repeat_region 18027..18705
/rpt_family="L1MB5"
repeat_region complement(18706..19168)
/rpt_family="MLT1B2"
repeat_region 19383..19447
/rpt_family="L2"
repeat_region complement(20015..20094)
/rpt_family="L2"
repeat_region 20108..20563
/rpt_family="MLT1F"
repeat_region 21080..21223
/rpt_family="L3"
repeat_region complement(21504..21725)
/rpt_family="MIR"
repeat_region complement(22444..22513)
/rpt_family="MIR3"
repeat_region 22970..23148
/rpt_family="G-rich"
repeat_region 24021..24120
/rpt_family="L2"
repeat_region complement(24730..24934)
/rpt_family="MIR"
repeat_region 24947..25068
/rpt_family="L2"
repeat_region 25137..25277
/rpt_family="MIR"
repeat_region 25441..25539
/rpt_family="MIR"
repeat_region 26018..26087
/rpt_family="L2"
repeat_region 27434..27568
/rpt_family="MIR"
repeat_region complement(28003..28114)
/rpt_family="MIR"
repeat_region 28115..28608
/rpt_family="MLT1H"
repeat_region complement(28622..29055)
/rpt_family="LMC1"

Query Match 25.8%; Score 515.2; DB 8; Length 115971;
Best Local Similarity 75.6%; Pred. No. 3.9e-71;
Matches 789; Conservative 0; Mismatches 203; Indels 52; Gaps 10;

QY 977 TACAGCATACCTTGAGAGTACTGTGGGTTGGTTCCATACACCAATATATACAAATA 1036
DB 84711 TACAGCATACCTCGGTATATTTGGGCTTGCTCCAGACCAAGCANTA----- 84761
QY 1037 TGCAAGAGTGATATCAATTAAGAGTACACAAAGTCTTTGGCTTCCAGTGCAT 1096
DB 84762 -----AAGTGAATATCAATTAAGAGTACATA--TTTTTGGTTTTCAGAGCAT 84813
QY 1097 ATAAAGTTTGCTTATCTACACTGTAGTCTGTTAAGTGTGCATATGTTATGCTTAA 1156
DB 84814 ATAAAGTTTATATCTATCTATCACTGATTAATTAATTTGCAATAGCATTTATGCTTAA 84873
QY 1157 AAAAACAATACCTTAATTTTAAATGCTTTATTAATAAATGCTTAACATCATTTGA 1216
DB 84874 AAATGTACATATCTTAATTTTAAATTAAGTATGCTTCAAAATGCTTAACATCATTA 84933
QY 1217 GCATTCAAGTGTAAATCTTTTGTCTGTGTGAAGTCTTTTCTTATTTGATGATGAT- 1275
DB 84934 GCCTTCAGTGAATCATATCTTTTGTCTGTGTGAAGAGGCTTGCCCTCAAGTGTGATGGCTG 84993
QY 1276 -----CGGGGTCAAGTGTCTGAAGCTTAAGGTGGCTGTGCAATTTCTTAA--- 1322
DB 84994 CCGATTGATCAAGCGGTGTGATGTGTGAAGGTGGGTGACATVAGCAATTTGTAAAGTA 85053
QY 1323 --ACAAAGTGAAGTTCAAATATCATGTTGACTCTTCTTTCATGAAGATTTCTCTCA 1380
DB 1323 --ACAAAGTGAAGTTCAAATATCATGTTGACTCTTCTTTCATGAAGATTTCTCTCA 1380
```



```

Db      85054 AGACAACAAAGAGTTGCTGCATCAATGACTCTCCCTTTCACAAAGATTTCTGTGA 85113
Qy      1381 GTGTGTGATGCTTTTGTATGACATTTTATGACAGTACAGTCTTTGAAATTTGA--TC 1439
Db      85114 GCATGTATGCTGTGTGACAGATTCATCCAGTATGAACTTTTCAAAATGAAATC 85173
Qy      1440 AATCTCTCAAAACCTGCTGCTTTTAAACAATTAATATAATTTGAAATTCAT 1499
Db      85174 AATCTCTCAAAACCTGCTGCTTTTATGACAGTATGAAATTTGAAATTCAT 85233
Qy      1500 TGTGTGATTTCAAAATTTTACAGTGTCTTACACAGAGTAAATTCATCTTCC 1559
Db      85234 TGTGTGATTTCAAAATTTTATGCAATCTTACAGAAATGATTCATCTCAAGAA 85293
Qy      1560 TGAGATGAAATCTTGTCTATCATCAATGAAATTTCTCTCTCTTCAAGTTTATCAT 1619
Db      85294 CCACTT-----TCTTGTCTTATTCATTAAGAGAACTCTCTCTCTTCAAGTTTATCAT 85349
Qy      1620 GAGATTTGACGAATACAGTCAATGCTTTCAGAGCTTCACTTCTTAAATTCAGTTCTCT 1679
Db      85350 GAGATTTGACGAATACAGTCAATCTTTCAGAGCTTTCAGAGCTTCTCTCTCTCTCTCT 85404
Qy      1680 TGTGTGATTTCAAAATTTTACAGTGTCTTACACAGAGTAAATTCATCTTCC 1739
Db      85405 TGTGTGATTTCAAAATTTTATGCAATCTTACAGAAATGATTCATCTCAAGAA 85464
Qy      1740 ATCCATGAGAGGTTGAAATGCACTTCTTCAAAATTCCTGTATATTTATTTGA---C 1796
Db      85465 ATTCACGAGAAATTTGAAATCACTCTTTTGAATCTCTGTATATGTTGCTTTTGAACCTC 85524
Qy      1797 CTCCCATGAAATCATGAAATGTTCTTAAATGACACCTTGAAATGTTGAATCTTTTCAAAAGGT 1856
Db      85525 CTCCCATGAAATCATGAAATGTTCTTAAATGACACCTTGAAATGTTGAATCTTTTCAAAAGGT 85583
Qy      1857 TTTCAATTTACTTACTTCCAGATTCATCCATCCAGAGATCCATTTCAATGCCATTTA 1916
Db      85584 TTTCAATTTACTTACTTCCAGATTCATCCATCCAGAGATCCATTTCAATGCCATTTA 85638
Qy      1917 GCCTTATGAAATGTTATTTCTTCAATTAATGAGCTTGAATTTGAATTTCTCTTGATCC 1976
Db      85639 GCCTTATGAAATGTTATTTCTTCAATTAATGAGCTTGAATTTGAATTTCTCTTGATCC 85698
Qy      1977 ATTTTCTGCAAAATGATGTTGTG 2000
Db      85699 ATGGGCTTCAGAAATGGGCTGTAG 85722

RESULT 24
AC124301/c 145713 bp DNA linear pri 30-AUG-2002
LOCUS AC124301 Homo sapiens chromosome 11, clone RP11-358H18, complete sequence.
DEFINITION AC124301
ACCESSION AC124301.6 GI:22549794
VERSION AC124301.6
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 145713)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-358H18
Unpublished
2 (bases 1 to 145713)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhassira,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelis,C., LaRoque,K.,

```

```

Lamazares,R., Landers,T., Lehoczkyl,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N.,
Mathews,C., McCarthy,M., McEwan,P., McKernan,K., Melidrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,T., Nguyen,C.,
Nicoll,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strause,N., Subramanian,A., Talamas,J., Testfaye,S.,
Theodore,J., Topham,K., Travers,M., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.,J.,
Young,G., Zainoun,D., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 145713)
REFERENCE
AUTHORS
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhassira,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kelis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Mathews,C.,
McCarthy,M., Melidrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,T., Nguyen,C., Nicoll,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,D.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 145713)
REFERENCE
AUTHORS
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhassira,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kelis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Mathews,C.,
McCarthy,M., Melidrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,T., Nguyen,C., Nicoll,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,D.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 30, 2002 this sequence version replaced gi:22507191.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL
COMMENT
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W18R
Web site: http://www-seq.wi.mit.edu
Contact: sequence submissions@genome.wi.mit.edu
Project Information
Center project name: L27330
Center clone name: 358_H_18

```



Db	4994	TGTTGATCTTCAACAAATGTTTATAGCATCTTCCACGAGAAATAGATTCATCTCAAGAA	4935
Qy	1560	TGAAATGGAATCTTTGCTCATATCCATTAAGAAGAAATTCCTCATCTGTTCAAATTTTATCAT	1619
Db	4934	CCACCTT-----TCTTGTCTTATTCATTAAGAACCACTCTCATCTGTTCAAATTTGATCAT	4879
Qy	1520	GAGATTGAGCAATATACAGTCATGCTTGAAGGCTTCACATTCACCTTTTATTTGACGTTCTT	1679
Db	4878	GAGATTGAGCAATATTCAGTCAATCTTGA-----CTTACTTCTTATTTTACTCTCT	4824
Qy	1580	TGCTGTTTCTTCAACAATCTGTGTTCTTCTTCCATTTGAAGTCTTGAAGCTTCCAAATC	1739
Db	4823	TGCTATTTTTATTCACATCTTGACAGTACTTCTCCACATGAGCTTGAAGCCATTAAAGTC	4764
Qy	1740	ATCCATGAGGTTGGAATTCGACTTCTTCCAAATTCCTGTATATTTATTTTGA---	1736
Db	4763	ATTACAGGAATTTGGAAATCAACTCTTTTGAACCTCTGTATATGTTGCTGTTTGAAGCTC	4704
Qy	1797	CTCCCATATATCATGATAGTGTCTTATATGAGCACTGGAATGTGTAATCTTTCCAAAAGT	1856
Db	4703	CTCCGATTAATTCACAAAGTCTTCTTAGTACACTTAGATGTGTAATTTCTTCCGAAGG-	4645
Qy	1857	TTTCAATTACTTACTGATCCAGATCCATCCATCCAGAGATCCACTTTCATGSCAGTTATA	1916
Db	4644	TTTCAATATGACTTTCGACAGATGTAT-----CAGAGGAATCAGTATCTATGGAAGACTTCA	4590
Qy	1917	GCTTATGGAATGTATTTCTTCAATAATAGCTTGAAGTGAATGAATTAATCTCTTGATCC	1976
Db	4589	GCTTATTAATAATGTATTTCTTAAAGAAATATGCTTGAAGTGAATTAATTAATCTCTTGATCC	4530

RESULT 25	HS168L15/c	178098 bp	DNA	linear	PRI 18-MAY-2005
LOCUS	HS168L15				
DEFINITION	Human DNA sequence from clone RPL168L15 on chromosome 6q26-27				

LOCUS	178098 bp	DNA	linear	PRI 18-MAY-2005
DEFINITION	HS168L15	Human DNA sequence from clone HS168L15 on chromosome 6p26-27		
		Contains the 3' end of the RPS6KA gene for ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK1), the gene for brain protein 44-1 like (BRP41), the 5' end of the gene for a novel protein similar to RIKEN cDNA 563040J11 (PRKRI) (MGC19825) and 5 CpG islands, complete sequence.		

ACCESSION	AL022069
VERSION	AL022069.1
KEYWORDS	HTG; BRP41; CpG island; MGC19825; pGR1; RPS6K2.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 178098)

REFERENCE 1 (bases 1 to 178098)  
AUTHORS Williams, S.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

COMMENT On Jun 26, 1998 this sequence replaced gi:2950404.

On Jun 26, 1998 this sequence version replaced gi:2950404.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone conligs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/MGP/Chr6>  
RP1-168J15 is from the library RPCI-1 constructed by the group of  
Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pCYPAC2

----- Genome Center -----

Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality  $\geq 30$ ); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

### source

gene

mRNA

CDS

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZBD:RRCIP704L15168"
/db_xref="taxon:9606"
/chromosome="6"
/map="q26-27"
/clone="RP1-168J15"
/clone.lib="RPIC1-1"
/join(2808049,1:25497, 25817,298049,1:113954, 114070,
298049,1:121425, 121506,6546, 6626,8621, 8700,
12281, 123397,15979, 16016,18233, 18395,26162, 26232,
27982, 28070,46398, 47062,57352, 57454,65670, 65800,
68054, 66179,84413, 84502,86232, 86450,93486, 93647,
96945, 97021,98561, 98678,102972, 103109,104016, 107517)
/gene="RPS6KA2"
/locus_tag="RP1-168J15.2-001"
/join(2808049,1:25497, 25817,298049,1:113954, 114070,
298049,1:121425, 121506,6546, 6626,8621, 8700,
12231, 123397,15979, 16016,18253, 18395,26162, 26232,
27982, 28070,46398, 47062,57352, 57454,65670, 65800,
68054, 66179,84413, 84502,86292, 86450,93486, 93647,
96945, 97021,98561, 98678,102972, 103109,104016, 107517)
/gene="RPS6KA2"
/locus_tag="RP1-168J15.2-001"
/product="ribosomal protein S6 kinase, 90kDa, polypeptide
2"
/notice="match: ESTs: AA409581 AA447506 AI903982 AJ453645
AL014466 AL517867 AL523835 AL523836 AU117630 AU130117
AU144411 AW410244 AW503268 AW654413 BE653300 BF205134
BF343539 BF513415 BE892517 BG295860 BG116876 BG490939
BI084035 BI198735 BI335924 BI836819 BM669400 BM699508
BM750594 BM766849 BM801851 BM873603 BQ029058 BQ180508
BQ277933 BQ303855 BQ421428 BQ430681 BQ431157 F06800 N79949
T74714 Z45429
match: CDNA5: AF140710 AJ131021 AK019881 BC002363 BC011189
BC014966 BC03013 L07597 L07598 X85106"
/join(298049,1:125719, 25817,298049,1:113954, 114070,
298049,1:121425, 121506,6546, 6626,8621, 8700,
12291, 123397,15979, 16016,18253, 18395,26162, 26232,
27982, 28070,46398, 47062,57352, 57454,65670, 65800,
68054, 66179,84413, 84502,86292, 86450,93486, 93647,
96945, 97021,98561, 98678,102972, 103109,104016, 107517)
/gene="RPS6KA2"
/locus_tag="RP1-168J15.2-001"
/standard_name="OTTHUMP0000017601"
/notice="match: proteins: BAB85907 P18653 P18654 P51812
O15349 Q21734 Q65331 Q96J38 Q9B829 Q9DC0 Q9PTN1 Q9WU73"
/codon_start=1
/product="ribosomal protein s6 kinase, 90kDa, polypeptide
2"
/protein_id="CA119651.1"
/db_xref="GI:56202585"
/db_xref="GOA:O15349"
/db_xref="UniProt/Swiss-Prots:O15349"
/translation="WMDSMKKFAARFSSVYLARKSKRSKSSLSLRLEEGVKEDIS
HHVGEFKPAPSOPEILKYLGQSSVYGLVLRKSGDGLYAKRVKATLKVRD
KRSRMEEDILAAVNNHPIYKILAAKPTETKYLILIDPLRGDILPRLSKVWPEED
KCVETLAETLAIDHLSIGITLYRDLPKRNILDLDBEGHIGITPDGLSKKAIIDDKKAS
KCEVETLAAPVNRKRGHTOSADWMSGVLMFEMTLQSLPFOCKDKETMALIYKAKL

```

[illegible]

	/db_xref="InterPro:IPR001245"	
	/db_xref="InterPro:IPR002290"	
	/db_xref="InterPro:IPR008271"	
	/db_xref="UniProt/TREMBL:Q5T801"	
	/translation="MPIADLEMLMKIEIEMPEIETTEEDLIADVEPTTEDPAEBSGS	
	DSAACTGVAGSEEEVGVKEIDISHNVKSGEKDPSPELLIKYLGQSYGVFLVR	
	KVKSDAQOLYAMKYLKKATLKLVRRVAVSKMERDLAEHNHFILKYHAPOTEGLL	
	LILDLRGDDLFTRLYSKEWTFLEDVKFLLALDALDHNSLGIITRLIKPENILLD	
	BEGRKINDFGLSKEAIDHDRAVSFCGTEYMAAEVNRHRTHOSADMFSGVLMFE	
	MLTGSLPFGDKRKETMALILKAKGMQPFSGEASVLRALFKENPCRLAGLDIGV	
	EIKRPFPFTIDANTLYRKEIKPPKFAVGDEPTHPDEPTKATPDSPGVPSA	
	NAHLFRGFSPVASTIQESOODLHKYPVHIYOOLGHNNHPFDGYIKXDIGGS	
	YSVCRCCHKATDIETAYAKIIDKSRDPSEEILIRGGQENIIITLKVDYDDKRVY	
	LVMEIMRGSELDRILRQRYSERSASVLCITYITMVDLISQSVNLRKSNLSILYR	
	DSEGSEIRVODFEGFAQLRANGMLMTWCPTTAIVPVYLKQCYDAACIMWSIGI	
	LLYTMLAGFTVPFANGDDTPPEILRIISGKALSGLGWMDSIDAKDVDASKLHYDP	
	HORLTMGLHPMVNREYLSNPQSLSDQVHLVYGAMAATYFALNRTPQARPLEEVL	
	SSNLARGRMKRILTSITRL"	

  

Query Match	25-7%	Score 514.4	DB 8	Length 178098	
Best Local Similarity	77.4%	Pred. No. 4.4e-71			
Matches 740	Conservative	0	Mismatches 176	Indels 40	Gaps 8
QY 1071	ACAAGCTTTTGCGCTCCCAAGTCGATATAAAGTTTGCTTATCTACTGTAGTCTGT	1130			
Db 167967	ATAAATTYYTGTYTCCACAGTACGTAAAAGTATATATACACTATCTATAGCTAT	167908			
QY 1131	TAAAGTGCANAGTGTATATGCTTAAAAA-----CACATACCTTATTTTAAANAAGC	1184			
Db 167907	TAAAGTGCATATAGCATTTATATGCTAAAAAACAAATGCTCATGCCCTTATTTAAAAATAC	167848			
QY 1185	TTTTATTACTAAAAATGCTAACATCATTTTGAGCATTCAGTAGTGTATCTTTTGCT	1244			
Db 167847	TTTTATTACTAAAAATGCTAGCAATCATCTGAACCTTCAGTAGTCATATATTTTGCT	167788			
QY 1245	GGTGAAGCTCTTTTCTTATGATGATCGAT-----CGGGCGTCAAGTCTGAA	1293			
Db 167787	AGTGAAGGCGCTTGCTGATGCTGATGGCTGCTGATCGATCAGATGAGTGTGTTGCTGAA	167728			
QY 1294	GCTTAAAGGCGGTGCGGCGATTTC-----TAAAACAAGTGAAGAATTGCAATATCAGT	1348			
Db 167727	GCTTAAAGGCGGTGCGGCGAATTTCTTAACATTAAGCAAGTGAAGTTTCTGCAATGAN	167668			
QY 1349	TGACTCTTCTTCATGAAAGATTCTCTCTAGTGTGATGCTTTTGATAGCATTTTA	1408			
Db 167667	TGACTCTTCTTCATGAAATATTTCTGATAGCATGTAATGCTGTTGATAGCATTTTA	167608			
QY 1409	TGCACAGTAAACCTTCTTGA AAAATTGA-TCAATCTCTCAAACCCTGCTGCTTTAA	1467			
Db 167607	CCCACAGTAAATTTATTTTCAAGAAATGGAAGTCAAGTCTCTTAAACCCTGCACTTTTA	167548			
QY 1468	CACCTAAGTTAATATAATATTCGAATCCATGTGTGTCATTTCACAAATTTTCAAGG	1527			
Db 167547	TTAATTAAGTTTACGCAATATTTCTAAATCTTGTGTGTCATTTCACAAATGTTTCAACA	167488			
QY 1528	TCTTACACAGAGTATGATTCATCTCATTTCTGATAGTGAATCTTGTGCTATCATAG	1587			
Db 167487	TCTTACACAGAGATGATTTCTGTCTCAGAAACCACTT---TCTTGTCTATTCGTAAG	167432			
QY 1588	AAGAAATTCCTCATCTGTTCAGATTATTCATGAGATGACAGCAATACAGTATGCTTC	1647			
Db 167431	AAGCAACCCCTCATCTGTTCAGATTTCATGAGATGACAGCAATACAGTATGCTTC	167372			
QY 1648	AGGCTCACTTCATCTTAAATTCAGGTTCTCTGTGTTTCTACACATCTGTGTTCT	1707			
Db 167371	AGG-----CTGATATTTTAAATTCAGGTTCTCTGTGTTTCTACACATCTGCAAGTACT	167317			
QY 1708	TCTTCATATGAAGTTTGAACCTCTCCAAGTCATCATGAGGGTTGGAATCATCTTCTC	1767			
Db 167316	TCTTCATATGAAGGTTGAACCCCTCAAAAGTCAACCAATGAGGGTTGGAATCAACTTCTC	167257			
QY 1768	CAATTCCTGTTAATTTATATTTTGAACCTCC---CATGAATCATGAATGTTCTTAATG	1824			

Db 167256 CAATTCCTGTTAAATGTTGATGTTTGAATGCTCCTCCATGATGATGATGTTTAAATG 167197  
QY 1825 GCACCTGGAATGGAATCCTTTCCAAAGGTTTCAATTACTTAATGTCAGATCCATCC 1884  
Db 167196 ACATCTAAATGATGCTCTTCCAGAAAGGTTTCAATTACTTCCCAATCAT-- 167139  
QY 1885 ATCCAGAGATCACTTTCATATGCGATGATATAGCTTATGAAATGATTTCTTCAATAT 1944  
Db 167138 ---CAGAGATCACTGCTATGACACTGATGCTTATTAAGATTTCTTCAATATAT 167082  
QY 1945 AAGCTTGAAGTGAATTAATCTCTTGAATCCTTTCTGCAAAATAGATGTTGG 2000  
Db 167081 ATGCTTGAAGTGAATGATCTCTTGAATCCTTGAATGATGATGATGATG 167026

RESULT 26  
AL357332 73991 bp DNA linear PRI 18-MAY-2005  
LOCUS Human DNA sequence from clone RP11-96C13 on chromosome 6, complete  
DEFINITION  
ACCESSION AL357332.11 GI:14270145  
VERSION HTG.  
KEYWORDS AL357332.11 GI:14270145  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homniidae; Homo.  
1 (bases 1 to 73991)  
Lovelil,J.  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
JOURNAL Cambridgehire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

COMMENT  
Clone request: clonerequest@sanger.ac.uk  
On May 31, 2001 this sequence version replaced gi:13157565.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
RP11-96C13 is from the library RP11-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk

-----  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.

FEATURES  
source  
1..73991  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-96C13"  
/clone\_1ib="RP11-11.1"  
100  
misc\_feature  
/note="Clone right\_end: RP1-224E15"  
11941..11972  
misc\_feature  
/note="Single clone region. Sequence from reads from a

misc\_feature  
short insert library derived from a single pUC clone.  
Restriction digest data confirm the assembly."  
73892  
/note="Clone\_left\_end: RP11-100A16"

ORIGIN  
Query Match 25.6%; Score 511.2; DB 8; Length 73891;  
Best Local Similarity 63.6%; Pred. No. 1.9e-70;  
Matches 1181; Conservative 0; Mismatches 563; Indels 112; Gaps 22;

Db 199 TTCTTCAGTCTTATGACCTCTTTTGTCTATTTGATGAGAGGCTTATATAGGGGC 258  
Db 70258 TCCTTCGATTAATTAAGCAGATGTTGCTTCAATATTTGAGAGCTTTTATTTATGATG 70317  
QY ACATCATTTAAATATGTTATGCTTATGATGATGATGATGATGATGATGATGATGATG 318  
Db 259 ATTTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70318  
QY ATTTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70318  
Db 319 ATTCATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 378  
Db 70374 TTTTATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70426  
QY ATAGGCACTG-GGCTTTTAAATAGTATTTTAAAGTATATATTTTCCCTTTT 437  
Db 70427 ATTAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70486  
QY TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 497  
Db 438 TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 497  
QY TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 497  
Db 70487 ACTTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70539  
QY TATATCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 557  
Db 498 TATATCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 557  
QY TACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70592  
Db 70540 TACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70592  
QY TAGGCACTTTCATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 617  
Db 558 TAGGCACTTTCATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 617  
QY CAGTCATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70652  
Db 70593 CAGTCATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70652  
QY TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 677  
Db 618 TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 677  
QY TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 677  
Db 70653 TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70712  
QY AGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 737  
Db 70713 AGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 737  
QY AAATTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 797  
Db 738 AAATTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 797  
QY CTCTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70829  
Db 70770 CTCTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70829  
QY ACCCTCAATAGTATTTTACAGAGCTCAAGTATGATGATGATGATGATGATGATGATGATG 857  
Db 798 ACCCTCAATAGTATTTTACAGAGCTCAAGTATGATGATGATGATGATGATGATGATGATG 857  
QY TCTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70888  
Db 70830 TCTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70888  
QY TCAATTTCTGCTCAATATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 917  
Db 858 TCAATTTCTGCTCAATATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 917  
QY CCAATTTTCTGCTCAATATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 917  
Db 70889 CCAATTTTCTGCTCAATATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 917  
QY TTT-----TTACTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 956  
Db 918 TTT-----TTACTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 956  
QY ATTAATTCACACAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 71008  
Db 70949 ATTAATTCACACAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 71008  
QY TTCTGTAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1016  
Db 957 TTCTGTAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1016  
QY TAAATTAATAAATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 71068  
Db 71009 TAAATTAATAAATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 71068  
QY CCACGATTAATAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1076  
Db 1017 CCACGATTAATAAATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1076  
QY CACGATTAATAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 71113  
Db 71069 CACGATTAATAAATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 71113  
QY CTTTGGCTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1136  
Db 1077 CTTTGGCTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1136  
QY TTTCTTG-TTTCCAGGCAATTAATAAGTATGATGATGATGATGATGATGATGATGATGATG 71171  
Db 71114 TTTCTTG-TTTCCAGGCAATTAATAAGTATGATGATGATGATGATGATGATGATGATGATG 71171

```

QY 1137 TCGAATAGTGTATGTCTAAAAAACA-----CATACCTTAATTTTAAATGCTTATTA 1191
Db 71172 TCGAATATCATATATGATATATTAACAATATGCAATCTTATTTGAAAAATCTTATG 71231
QY 1192 CTAAAAAATGCTAAACATCATTTAGACATTCAGTGTGTAATCTTTTGTGCTGAGAA 1251
Db 71232 CTAAAAATACCTAACCATAGTGTGAGTTTCACTAGTGTATATCTTTTGTGCTGAGAG 71291
QY 1252 GGTCTTTTCTTATGATGCTGAT-----CGGGGGTCAAGGTCTGAAGCTTAg 1300
Db 71292 GGTCTTCTGCTGATGATGCTGCTAACTCATCAGATGCTGATGCTGAAGCTTAT 71351
QY 1301 GTGGCTGTGGCACTTTCTTAAACAACA-----GTGAAGATTGAATATCAGTGAAGCTT 1356
Db 71352 GCAGATGTGCCAATTTCTTAAATTAAGAAAAATGATGTTCCACGGTGATGACTCTT 71411
QY 1357 CCTTTCATGAAAAGATTTCTCTAGTGTGATGCTTTTGTATGATGATTTATGACAGT 1416
Db 71412 TCTTTTCATGAAAAGTTTCTCTGTAGCATGCAACATTTTGTATGTTTCCCCACAGT 71471
QY 1417 AGAATCTTTTGAATAATGGA-TCGAATCTCTCAAAACCTGCTCTGCTTTTACACCTTA 1475
Db 71472 AGAATCTTTTCAAAATTTAGATCAATCTCTAAACTGCTGTTGTTTATCAGCTAA 71531
QY 1476 GTTAATATATATTTCTGAAATCCATTTGTGTATTTTCAACAATTTTCAAGTCTTAC 1535
Db 71532 GTTATATATATTTCTTAAATTAATCTGTTGTATTTCAACAATTTTCAAGCATCTTCA 71591
QY 1536 AGGAGTAGATTCATCTCATTTCTGATGATGAATCTTTGCTATCATCAATGAAGAAT 1595
Db 71592 AGGAGCATTTTCATCATCATACATACCACTT---TCTTGTCTATTCATTAAGGGCACT 71647
QY 1596 CCTCATCTGTCTCAAGTTTATCATGAGATTTGACAGCAATACATGCTTCTCAGGCTCA 1655
Db 71648 CCTCATCTGTCTCAAGTTTATCATGAGATTTAAAGCAATTCAGACATCTTCAAGG--- 71703
QY 1656 CTTCACCTTTAATTTCAAGTCTCTGTGCTTTTACACATCTGTGCTTCTCTCAT 1715
Db 71704 ---CACTCAAAATTAATATGTTCTCTTATTTTCCACACATTTGTATTAATCTCTCAT 71760
QY 1716 TGAAGCTTTGAA-CCTCTCCAAATCCATCCATGAGGGTGAATGCACTTCTCCAAATTC 1774
Db 71761 GGAAGCTTTGAACTCTCTCAAAATTAATTAAGGGTGAATCAACTCTTCTTAAATTC 71820
QY 1775 CTGTATAATTTTATTTATTTGACCTC-CCATGAATCATGAATGTTCTTAATGCACTGA 1833
Db 71821 CTATTAATGTTCAATTTCTCACTCTTCCCAATCATGAATGTTGCAATGGCATTGA 71880
QY 1834 ATGCTGAAT-CTTTTCCAAAAGTTTCAATTTACTTAATGCTCCAGATCCATCCAGAG 1892
Db 71881 ATATTAATATCCCTTTTCAAGGGTGTTCATTTACTTGTCTCAGATCCATCAAGAAAC 71940
QY 1893 GAT-----CCACTTCAATGCCAGTTATAGCCTTAATGAGATGATTTCTTCATTAAT 1944
Db 71941 ACTATCTAATGAGAGCTATGATGCTTCAATATCATTTAATAATGATTTCTTAATAAT 72000
QY 1945 AAGGCTTGAAGTTGAATTAATCTCTGATTCATTTTTCGCAAAATAGATGTTGTG 2000
Db 72001 AAGACTT-AAAGTCAAAATGACCTCTTGAATCCATGGGCTGCAATGATGATGTTGTG 72055

```

```

RESULT 27
AL590674/c AL590674 114056 bp DNA linear PRI 24-OCT-2001
LOCUS Human DNA sequence from clone RP11-286N24 on chromosome 1, complete
DEFINITION
sequence.
ACCESSION AL590674.5 GI:16444743
VERSION AL590674.5
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
1 (bases 1 to 114056)
AUTHORS
Almeida, J.
TITLE
Direct Submision
JOURNAL
Submitted (24-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
request: clonerequest@sanger.ac.uk
COMMENT
On Oct 25, 2001 this sequence version replaced gi:13990635.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-286N24 is from the library RPl-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-286N24. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP11-286N24 is at 114056 in this
sequence. The true left end of clone RP11-2113 is at 724 in this
sequence. The true right end of clone RP11-152M20 is at 2000 in
this sequence.
FEATURES
Location/Qualifiers
source
1..114056
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-286N24"
/clone_1ib="RPl-11.1"
6238..6251
/misc_feature
/note="Sequence from overlapping clone ba2L313 (AL596266).
Assembly confirmed by restriction digest."
7035..7250
/misc_feature
/note="Sequence from overlapping clone ba2L313 (AL596266).
Assembly confirmed by restriction digest."
ORIGIN
Query Match 25.6%; Score 511.2; DB 8; Length 114056;
Best Local Similarity 58.3%; Pred. No. 1.6e-70;
Matches 1218; Conservative 0; Mismatches 758; Indels 112; Gaps 14;
QY 3 ATTCTCTGCTTCAGCTTCCCAAGTAGTGAAGTTACAGTGTCCAGCCACCAAGTCCAGC 62
Db 94214 ATTCTCTGCTTCAGCTTCTAGTAGTGTGAGATTAAGAGCAACCCGCAATGCCCCAGC 94155
QY 63 TAAATTTTGA--TTTATAGAGAGCGGGCTTACCGGTTTGCAGGCTGTATCAAA 120
Db 94154 TAAATTTTGAATTTTATAGAGAGTGGATTTTACATATGTGGCCAGGCTGTCTCAAA 94095
QY 121 CTCCTGACCTCAAGTAGATGCTGCTGCTCAGCTCCCAAAATGCTAGATTAAGATGTG 180
Db 94094 CTCCTGACCTCAAGTAGATTAAGATGCTGCTCAGCTCCCAAAATGCTGAGATTAAGCATG 94035

```



181 AGTCAACCCACCCACCCCTTCTTCAAGTCTATCACTCTTTTGTCTA--TATTGTG 238  
DB AGCTACAGTGGCCAGCTTACATCATTTATTAGCTTTTATAATAGGAATAATGTT 93975  
QY 239 AGAGCTTATTATTAGGGGACATACATTAAATTTGTAATCTTATTGATGATGAT 298  
DB 93974 TATGTTTAAATTAATAAATTAATCACTAATAAATATGTTTCCATTTATGTTAAACC 93915  
QY 299 CTGATATTATGATGTCGTATTCATTCCTCCGATAGTATTTCTTTTCTAAATTTTTC 358  
DB 93914 TATGATGTGCAAAAAAATTTTACATTAATAAATATATGATGATCTTTTATGATGACA 93855  
QY 359 TGAATGTCTGCTATTAAACATAGCCACTGTGGCTTTTAAATAGTATTTATGTA 418  
DB 93854 GGATTTATAGCTAATTTTACTTGTGATTAATCAATCCAAATGTTTCACTTTTCAATGTT 93795  
QY 419 TATATTTTCTTT--TTTTTTTTTAAAGTTTAAAGTTTAAAGTTTCTTAACTTAA 476  
DB 93794 CTTTATTTGTTGGTTTGTGTTGTTAGTTTAAAGTTTAAATTTATATTAATTTAA 93735  
QY 477 GTGGGATCTTATAGGACATATAATCTGGGCTTGAAGTATTTATTTATCTGATATCT 536  
DB 93734 TCTCTGTAACATGATTAATTTGCAATATTTTCTGTTCAATGATGTTTCTTCT 93675  
QY 537 CAACCTTTTGTGAGTGTATAGGCACTTACATTTAGTGAATTTATAGACATGTTTG 596  
DB 93674 CTTTCTGTATAGTACATTTATGACAAAGTTCTTAAATTTGATGAGTCCAAATGAT 93615  
QY 597 ATTGCTATACACTCTTTCAATTTGTTTATATGAGCCATCTTCAATGTTCTTTT 656  
DB 93614 AATATGTTTCTTTGTTGCTGATATTCGGTGTGATCCAGATCCAAATATATGCCCMAAT 93555  
QY 657 TCATCTTTGACATTTCTTTAGTACATACTTTTGTATTTATTTATATATCTATG 716  
DB 93554 CCAAGTCAVAAAAGAT--TTTCCCAATGACCTTCAATTTTGAAGTATTTTCTG 93497  
QY 717 GCTTTTATGTTATACCTCTTAAATTTTCTTTTCTTTTATAGTATTTT--TATTAAT 772  
DB 93496 GATATTTGATTTCTAGTGTCTTCCACTGCTGTAATTTATGATTTGATGTTGCTATGAC 93437  
QY 773 ATACATCTTTAACTTATCAAGATTAACCTTCAATAGTATTTTCAAGCTCAATGTAAT 832  
DB 93436 TTGCATTTATTTAATAGAGAAGTGAATATTTTATCTTTGTTCTCCATGTTGTAAT 93377  
QY 833 GTAGAAACCTTACAAAGATTAATTTCTATCTGTCTCTAATTTTATGCTATGCTAAT 892  
DB 93376 GTCTTGTCTGTCTTTAAGATATTTCTTATCACTGTGCTTCAAGCAATTTGATTAATA 93317  
QY 893 AATACATAGTGTGTTGTTGTTGTTTATACCTTATGCTGTGCTGGGCTCAGCAAA 952  
DB 93316 TATACCTTGTGTGTTTGTGTTGTTTATACCTGTGAGTTCATGATTTTGTAG 93257  
QY 953 CATTTTCTGTAAGGGCTAGATATGTAAGGATATC-----CTTGA 993  
DB 93256 CCTGTGATTTTATGATTCATCAAAATTTGAAAACATTTCTTCATTTATTCATATACA 93197  
QY 994 GATACGTGTGGTTGGTTCCATACCAACATTAATACAAATATGCAAG----- 1042  
DB 93196 TTTTCTTTGTCTTCTCTCTTTCTTAATTAATAGGGGATACCCAGAGATATTTGTG 93137  
QY 1043 -----AAGTATATCAATTAAGTATGACACAAAGTC 1077  
DB 93136 GGTTCATTTATGACACATGCAATAAAGTATTCATTAAGCAATATCACACAACT 93077  
QY 1078 TTTTGGCTTCCAGTCAATTAAGTTTGTATATCACTGTAGTCTGTTAAAGT 1137  
DB 93076 TTTTGGTTCCAGTCAATTAAGTTTGTATATGTTTACATTAATATAGTCTATTAAGTGT 93017  
QY 1138 GCAATGTTATGTTCTTAATAAAGC-----ATACCTTAATTTTAAATGCTTATTAATCT 1193  
DB 93016 AATATAGTATTAATGCTTAATAAAGTATGATTAATATCTTAATGAATAAATACCTTATTTGCT 92957

QY 1194 AAAAAATCTTAACAATCAATTGAGCATTCAGTGTGTAATCTTTTGTGCTGTGAAAG 1253  
DB 92956 AAAATATGCTTAAGATTAATTCAGAGCTTCAAGAAAGTCATTAATCTTTTGTGCTGTGAAAG 92897  
QY 1254 TCTTTTCTTATGATGATG-----ATGGGGGGTCAAGGCTGAAGCTTAAAGGT 1302  
DB 92896 TCTTGTCTCAATGTTGATGGCACTGATGATCAGGGGATGAGCTGTGAGAGTTTGGGT 92837  
QY 1303 GCGTGTGACAGTTCTTAAACAAACAGTGAATTTGATATCA-----GTGACTCTTCC 1358  
DB 92836 GCGTGTGACAGTTTAAATAATTAAGCAAAATATGAAGTTTCCACATGTTGATCTTTC 92777  
QY 1359 TTTCAVGAAGATTTCTCTAGTGTGATGCTTTTGTATGACATTTTATGACACATG 1418  
DB 92776 TTTCAACGTATGATTTCTGTATGATVGAAGATGCTGTGATATGATTTTATCACTAATG 92717  
QY 1419 AACTCTTTGAAAAATTTGATCAATCCCTGCAACCTGCTGCTTTTAAACAACCTAAGT 1478  
DB 92716 AACTCTTTGATTTATGATGATCA--ATTCACAAACCTGCTGCTTTATCAACTAAGT 92658  
QY 1479 AATATATATTTCTGAATCCATTTGTTGATTTCAACAAATTTTCAAGTCTTTCACAG 1538  
DB 92657 TATGATATCTTAAATCTTTGTTATTTCAACAGTGTTCACAGCTTTTCCAGA 92598  
QY 1539 AGTATATTCATCTGATTTCTGAGATGGAATCTTTGCTCAT-----CAATTAAGAA 1592  
DB 92597 AGTATATTTCTTCAAGAAACCTTAAGACCTTTGCTTCCATATGATACATTAAGAAAGC 92538  
QY 1593 ATTCTCATCTGTCAAGTTTATGATGATGAGAAATGAGCAATGCTGATGCTTCAAGGC 1652  
DB 92537 ATTCTCATCTCAATCAAAATTTATCAATGATGAGCAATTCAGTCTCATCTTCAAGGC 92479  
QY 1653 TCACCTTCACTTTAATTCAGTCTCTGCTGTTTCTACACATCTGTGTTCTTCTCTC 1712  
DB 92478 -----TCTACTTCTAGTCTCTGCTGTTTCTCAACATTTGAGAGGCTTCTC 92429  
QY 1713 CATTAAGCTTGAACCTCTCCAGTCAATCCATGAGGGTGAATGATCTTCTCAAT 1772  
DB 92428 CACTGAGGCTTGAACCCCTCAAGTCAATCCAGAGATGGAATCAATCTTCTTCAAC 92369  
QY 1773 TCCGTATATTTATTTATTTGACCTCCAGTAATCAATGAATGTTCTTAATGAGCACTG 1832  
DB 92368 TCCCATTAATGTTGATCTTTGATGCTCATTAATCAAGATGTTCTTAATGACATG 92309  
QY 1833 AATGATGATCTTCTTCAAAAGTGTTCATTTACTTATGATCCAGATCCATCCAGAG 1892  
DB 92308 AATGATGATCTTCTTCAAGAGTGTTCATTTACTTATGATCCATC-----A 92257  
QY 1893 GATCACTTTCAGTCAATGACCTTAATGACCTTAATGAAATGATTTCTTCAATTAATGAGCTTG 1952  
DB 92256 GAGAAATCACTATGCAACTAATATGCTTACAGCATGATTTCTTAATTAATGAATTTG 92197  
QY 1953 AAAATTTGAATTAATCTTCAATGATCCATTTTCTGCAAAATGATGTTGTG 2000  
DB 92196 ATAGTTGAATTAATCTTCAATGATGAGTTTCAAGATGATTTGTG 92149

RESULT 28  
AL596266/c 128479 bp DNA linear PRI 18-MAY-2005  
LOCUS Human DNA sequence from clone RP11-2L3 on chromosome 1 Contains a  
DEFINITION translational inhibitor protein p14.5 (UK14) pseudogene, complete  
sequence.  
ACCESSION AL596266 GI:15626484  
VERSION AL596266.3  
KEYWORDS HTG; UK14.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 128479)  
AUTHORS Chapman,J.



TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

## COMMENT

Clone requests: clonerequests@sanger.ac.uk  
On Sep 16, 2001 this sequence version replaced gi:15028804.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP11-2L13 is from the library RPL1-11.1 constructed by the group of  
Pleier de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.

## FEATURES

source Location/Qualifiers

1..128479  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"

/clone="RP11-2L13"  
/clone\_1ib="RPCI-11.1"

misc\_feature

/note="Clone left\_end: RP11-2L13"

gene

75806..76205

/locus\_tag="RP11-2L13.1-001"

CDS

75806..76205

/locus\_tag="RP11-2L13.1-001"

/note="match: proteins: SW:P52758 SW:P52759 SW:P52760  
Sw:P80601 Tr:O97509 Tr:Q9MUW8"

/pseudo

/codon\_start=1

/product="translational inhibitor protein p14.5 (UK114)  
pseudogene"

misc\_feature

/note="Clone right\_end: RP11-286N24"

misc\_feature

126480  
/note="Clone left\_end: RP11-567B20"

## ORIGIN

Query Match 25.6%; Score 511.2; DB 8; Length 128479;  
Best Local Similarity 58.3%; Pred. No. 1.6e-70;  
Matches 1218; Conservative 0; Mismatches 758; Indels 112; Gaps 14;  
QY 3 ATTCTCTGCTCAGGCTCCCAAGTAGCTGAGGTACAGGTGCGACGCCACACCTCCACG 62  
DB 93497 ATTCTCTGCTCAGGCTCTTCTGAGTAGCTGGGATTCAGGCACCCGCCATCATGCCACG 93438  
QY 63 TAAATTTTGTAA--TTTAAAGAGACGGGGTTTCAACGTTTGGCCAGGCTGGTATCAAA 120  
DB 93437 TAAATTTTGTATTTTAAAGAGATGGGATTTTCAACATGTTGGCCAGGCTGGTCTCAAA 93378  
QY 121 CTCCTGACCTCAAGATGATCTGCTGCTCAGCCTCCCAAAATGCTAGATTACAGGTGTG 180  
DB 93377 CTCCTGACCTCAAGATGATCAACCATCTCAGCCTCCCAAAAGTCTGGATTTATACCATG 93318

```

OY 1194 AAAAAATCTAACATCATTTGACATTCAGTGTGTAATCTTTTGGCTGGGAAdG 1253
Db 92239 AAAAAATCTAACATTCATTCAGAGCTTCAGAAAGCATTAATCTTTTGGCTGG393AG 92180
OY 1254 TCTTTCTTAATGAGACG-----ATGGGGGGCTCAGGGGCTGAAGCTTAAGGGT 1302
Db 92179 TCTTTCTTAATGAGAGGCACTGATCAGTACAGGATGAGTGGCTGGAGGTTTGGGT 92120
OY 1303 GCGCTGGCAGTCTTCTTAAACAGAGTAAGATTCATATCA-----GTTGACTCTTCC 1358
Db 92119 GCGCTGGCAGTCTTCTTAAACAGAGTAAGATTCATATCAAGTTCAGTCTTCTTCC 92060
OY 1359 TTTCAAGAAAGATTCTCTCTAGTGTGATGCTTTTGTATGACATTTTATGCAAGTNG 1418
Db 92059 TTTCAAGAAAGATTCTCTCTAGTGTGATGCTTTTGTATGACATTTTATGCAAGTNG 92000
OY 1419 AACTCTTGAAGATTTGATTCATATCCTCCCAACCTGCTGCTTAACTTAACCTTAAGTT 1478
Db 91999 AACTCTTGAAGATTTGATTCATATCCTCCCAACCTGCTGCTTAACTTAAGTTG 91941
OY 1479 AATATATATCTGATTCATTCATTTGTCATTTTCAACAATTTTCAAGTGTCTTCAACGAG 1538
Db 91940 TATGAGATCTTCTTAATCTTTGTTGTTATTTTCAACAGTGTTCAGAGCTTTTCCGANA 91881
OY 1539 AGTATATTCATCTCATTTCTCGAGATGGAATCTTGTCTCAT-----CCATTAAGAA 1592
Db 91880 AGTATATTCATCTCATTTCTCGAGATGGAATCTTGTCTCAT-----CCATTAAGAA 91821
OY 1593 ATTCTCATCTGTCATGATTTATGATGATTTGAGCAATACAGTACATGCTTCAAGGCC 1652
Db 91820 ATTCTCATCTGTCATGATTTATGATGATTTGAGCAATACAGTACATGCTTCAAGGCC 91762
OY 1653 TCACCTCACTTTTAAATTCAGATCTCTGCTGTTTCTTACACATCTGCTGTTCTTCTC 1712
Db 91761 -----TCTACTCTAGTCTCTGCTGTTTCTTACACATCTGCTGTTCTTCTC 91712
OY 1713 CATGGAAGTCTTGAAGCTTCTCCAGTATCATGATGAGGTTTGAATGACTTCTTCCAAT 1772
Db 91711 CACTGAGGCTTGAAGCCCTCCAAAGCATTCACAGAGATTTGGAATCACTTCTTCAAA 91652
OY 1773 TCTCTGTAATTTATTTTGAACCTCCCATGAAATCATGAACTTTTAAATGGCACTTGG 1832
Db 91651 TCCCATTAATGTTGATCTTTTATGCTGCTATGATCATCAGATGTTCTTATATGACATCTG 91592
OY 1833 AATGAGATCTTCTTCAAAAGTCTTCAATTTACTTAACTTCAAGTATCATCATCAGAG 1892
Db 91591 AATGAGATCTTCTTCAAAAGTCTTCAATTTACTTAACTTCAAGTATCATCATCAGAG 91540
OY 1893 GATCCACTTCAATGCGCAGTTATAGCCTTATGGAATGTAATTTCTCAATPATTAAGGCTTG 1952
Db 91539 GAGAAATCACTATGCGCACTAATATAGCTTACAGCATGTAATTTCTAATATTAAGATTTG 91480
OY 1953 AAAAGTGAATTTACTCTTGTATTCATTTCTGCAAAATAGATGTTGTG 2000
Db 91479 ATAGTTGAATTAATCTCTTGTATTCATGAGTTACAGATGATTTGTG 91432

```

```

RESULT 29
HSDJ45N11/c 99025 bp DNA linear PRI 18-MAY-2005
LOCUS Human DNA sequence from clone RPI-45N11 on chromosome 6q15-16.1,
DEFINITION complete sequence.
ACCESSION AL132996
VERSION AL132996.4 GI:8648910
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 99025)
AUTHORS Collier,R.
TITLE Direct Submission

```

## JOURNAL

## COMMENT

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
 Clone requests: clonerequest@sanger.ac.uk  
 On Jun 21, 2000 this sequence version replaced gi:7940170.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Bm, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMBEP; Information on the WORMBEP database can be found at:  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormbep](http://www.sanger.ac.uk/Projects/C_elegans/wormbep)  
 This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at:  
<http://www.sanger.ac.uk/Chp/Chr6>  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: vegas@sanger.ac.uk  
 -----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 RPI-45N11 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pCYPAC2.

## FEATURES

## source

location/Qualifiers

1..99025

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="RZPD:RPI704N1145"

/db\_xref="taxon:9606"

/chromosome="6"

/map="q15-16.1"

/clone="RPI-45N11"

/clone\_lib="RPI-1"

1/note="clone\_left\_end: RPI-45N11"

99025/note="clone\_right\_end: RPI-45N11"

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

```

ORIGIN
Query Match 25.6%; Score 511; DB 8; Length 99025;
Best Local Similarity 75.2%; Pred. No. 1.8e-70;
Matches 786; Conservative 0; Mismatches 210; Indels 49; Gaps 10;
975 AGTACAGCATCTTGGAGATACCTGAGTTGGTTCATACCAACCAATTAATACAA 1034
75033 AGTACAGCATCTTGGAGATACCTGAGTTGGTTCATACCAACCAATTAATACAA 74981
1035 TATGACAGAGTATATACATTAAGTATGATCACAAGCTTTTGGCTCCAGTGC 1094
74980 -----AAATGAACATCACATTAAGCCAGTACGTAATTTGTTTCCAGTTC 74929
1095 ATATTAAGATTTTGTATATACACCTGATCTGTTAAGTGAATA--GTTATATG 1152
74928 ATATTAAGATTTTGTATATACACCTGATCTGTTAAGTGAATA--GTTATATG 74869
1153 CTAAAAAACAATACCTTAATTTTAAATGCTTTTAACTTAATAAATCTTAACATCAT 1212
74868 TTTAAAAATGATACCTTAATTTTAAATGCTTTTAACTTAATAAATCTTAACATCAT 74809
1213 TTGAGCATTCAGTATGTGAA--TCTTTTGTCTGTGGAAGGCTTTTCTTATTTGATGA 1270
74808 CTGAGCTTCAGTATGTGAA--TCTTTTGTCTGTGGAAGGCTTTTCTTATTTGATGA 74749
1271 CTGAT-----CGGGGCTCAGTCTGATGAGTTGAGGCTGCTGGCAGTTCTT 1319
74748 TGGCTGCTGACTCATCAGGCTGTGTGCTGAAGGTTGGAGTGTGTGGCACTTCTT 74689

```

OY	1320	AAAAACAGGAAAGATGTGCAATTCATTCAGTTGACCTTCCTTATGAAAGATTTCTCT	1379
Db	74688	AAAATTCAGA-AAAATTAATTAATGATTCATATGACTCTTTCTTATGAAAGATTTCACTGT	746390
OY	1380	AGTGTGTGATGCTTTTGTATGACATTTTATGACAGTAGAATCTTTGAAATTTGGA-T	1438
Db	74629	AGCATGCAATCTGTATTGACAGCATTTATACCAGTAAATCTTTTCAAAATTTGTACT	74570
OY	1439	CAATTCCTTCAACCCCTGCTCTGCTTTAACAACCTAAGTTAAATATATATTTTGAATCCA	1498
Db	74569	CAATTCCTTCAACCCCTGCTCTGCTTTAACAAGTAAATGTTATATGTATTTCAACTCCCT	74510
OY	1499	TTGTGTTCATTTCAACAATTTTTCACAGTGTCTTCACACAGGAGTAGATTCATCTCATTTTC	1558
Db	74509	TTGTGTTCATTTCAACAACATTTTCACACACCTTCATAGAGTAGATTCATCTCAAGAA	74450
OY	1559	CTGAGATGGAATCTTGTGCTCATCCATTAAGAAGAAATTCATCTGTCTCAAGTTTATCA	1618
Db	74449	ACCAACT---TCTTTACTATCCATTAAGAAGAACTCCATCTGTTCAAACTTATCT	74394
OY	1619	TGAATTTGACACATTAACATGATCTCTTCACAGGCTCACTTCACCTTTTAATTCAGTTTTC	1678
Db	74393	TGAATTTGACACAAATTCCTGACCTCTTCAGG-----CTTCACCTTGAAATTTCACTTATTC	74339
OY	1679	TTGCTGTTTTCACACACATCTGTGGTTCCTTCCTTCATTTGAAGCTTGAACCTCCCAACT	1738
Db	74338	TTGCTATTTTCATCATCACACTGCACTTACTCTTCACATTTAGATTTGAACCCCTCAAACT	74279
OY	1739	CATCATGAGGGTTGGAATGCACTTCTTCCAAATTCCTGTTAAATTTATATTTTGAAC--	1796
Db	74278	CATTCATGAGGGTTGGAATCAACTTCTTCATCTCTCTGTTAAGTTATATTTGACCT	74219
OY	1797	-CTCCCATGGAATCATGAAATGTTCTTATATGGCACTGGAATGTGGAATCTTTTCCAAAAG	1855
Db	74218	GCTCCCATGGAATCAACAATGTTCTTAAATGGCACTTGAAGAAGAAATTTCTTCCGGAAG	74159
OY	1856	TTTTCAATTTACTAGTCCAGATCCATCCATCCACAGAGGATCCACTTCATGCGCACTTAT	1915
Db	74158	TTTTCAATTTACTTGTCCAGATCCAT-----TAGAGGAATCACTATCTATAGGCACTTAT	74104
OY	1916	AGCCTTATGGAATGATTTTCTTCATATTAAGGCTTGAAGTTGAATTAATCTCTTGAATC	1975
Db	74103	AGCTTTTGAAGATGTATTTCTTAAATAGAAGCTTGAAGTTGAATTAATCTCTTGGTT	74044
OY	1976	CATTTCTCGAATAATGATGTTGG 2000	
Db	74043	CATGGGCTCGAAGATGATGTTGG 74019	
RESULT 30			
AC027156			
LOCUS			
DEFINITION			
AC027156			
ACCESSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
AUTHORS			

Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, M., Gage, D., Gajagan, J., Gardyas, S., Ginde, S., Goylete, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Klein, J., LaRoque, K., Lamazares, R., Jones, C., Kann, L., Karatas, A., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 172061)

Bitren, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barua, N., Baetien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campioiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Gajagan, J., Gardyas, S., Ginde, S., Goylete, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howard, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Lander, E., Lehotko, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (24-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 19, 2000 this sequence version replaced gi:7331458.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center clone name: L5972  
Center project name: 450 D 8  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 161486 bases at least Q40  
Consensus quality: 167277 bases at least Q30  
Consensus quality: 169194 bases at least Q20  
Insert size: 178000; agarose-fp  
Insert size: 170261; sum-of-contigs  
Quality coverage: 4.3 in Q20 bases; agarose-fp  
Quality coverage: 4.5 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown.



OY		1499	TGTTGGTCATTTCAACAATTTTTCAAGTGTCTTCCACAGAGAATGCATCTCATTTTC	1558
Dd		86188	TTGTTGTTCATTTCACAACAATTCACAGCACCCTTCATAGAGTAGAATTCATCTCAAGAA	86244
OY		1559	CTGAATGGAATCTTGTCCTCATCCATAAGAAATTTCTCATCTGTCTCAAGTTTATCA	1618
Dd		86248	ACCACT-----TCTTTACTATCCATTAAGAAAGAACCTCATCTGTTCAACTTTATCG	86303
OY		1619	TGAATTTGCAGCATAATACATGTCATCTTCAGGCTCACCTTCACCTTTTAATCCAGTTTC	1678
Dd		86304	TGAATATGCAGCAATTTCTCGACATCTTCAGG-----CTTCACCTTGGAAATTCATCTTATC	86358
OY		1679	TTGCTGTTTTCAACACATCTGTGGTTCCTTCCTCCATGAAGCTTGAACCTGCCAAGT	1738
Dd		86359	TTGCTATTTTCATACACAATCGCAGTTACTCTTCACCTTAGATTTGAACCCCTCAAGT	86418
OY		1739	CATCCATGAGGGTTGGAATGCATCTTCCCAAATTCCTGTTAATATTTATATTTTGAC--	1796
Dd		86419	CATTCATGAGGTTTGGAAATCAACTTCTCCATCTCCTGTAAGTTTATTTTACCT	86478
OY		1797	-CTCCCATGAAATCATGAATGTTCTTAATGCGACCTGGAATGTGAATCCTTTCCAAAAGS	1855
Dd		86479	GCTCCCATGAAATCAACAATGTTCTTAATGCACTCAACAAGAAATTCCTTCCGGAAGS	86538
OY		1856	TTTTCAATTTACTAGTCCAGATCCATCCATCCAGAGATCACTTCAATGCCAGATTAT	1915
Dd		86539	TTTTCAATTTACTTGGCCAGATCCAT-----TAGAGAAATCACTATCTTAGGCACTAT	86599
OY		1916	AGCCTTAATGGAATGATTTCTTCATATATAAGGCTTGAAAGTTGAATTAATCTCTTAATC	1975
Dd		86594	AGCTTATATGAATGTATTTCTTAATAAGAAGCTTGAAAGTTGAATTAATCTTTGGTT	86653
OY		1976	CATTTTCTGCANAATAGATGTTGCG	2000
Dd		86654	CATGGGCTGCAGATAGATGTTGCG	86678
RESULT_31				
HSJ37OM22				
LOCUS		HSJ37OM22	143747 bp	DNA linear PRI 18-MAY-2005
DEFINITION		Human DNA sequence from clone RP3-37OM22 on chromosome 22, complete		
ACCESSION		282206		
VERSION		282206.1		
KEYWORDS		HTG.		
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
		Homidae; Homo.		
		1 (bases 1 to 143747)		
REFERENCE		Burgess,J.		
AUTHORS		Direct Submission		
TITLE		Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,		
JOURNAL		Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk		
COMMENT		Clone requests: clonerequests@sanger.ac.uk		
		On Dec 12, 1997 this sequence version replaced gi:1772957.		
		The following abbreviations are used to associate primary accession		
		numbers given in the feature table with their source databases:		
		Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WormPep; Information		
		on the WormPep database can be found at		
		http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence		
		was generated from part of bacterial clone contigs of human 22		
		chromosome 22, constructed by the Sanger Centre Chromosome 22		
		Mapping Group. Further information can be found at		
		http://www.sanger.ac.uk/HGP/Chr22		
		RP3-37OM22 is from the library RPCI-3 constructed by the group of		
		Pietier de Jong. For further details see		
		http://www.chori.org/bacpac/home.htm		
		VECTOR: pCYRAC2		
		----- Genome Center		
		Center: Wellcome Trust Sanger Institute		

```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest
except on the rare occasion of the clone being a YAC.

    location/Qualifiers
1..143747

```

```

UNES
source
    Location/Qualifiers
    1..143747
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="R2PD:R2PCLP704M22370"
    /db_xref="taxon:9606"
    /chromosome="22"
    /clone="RP3-370M22"
    /clone_1fb="RPCL-3"
    complement(join(3536..4154,7268..7510,32902..33405,
    39197..39292))
    /locus_tag="RP3-370M22.3-003"
    complement(join(3536..4154,7268..7510,32902..33405,
    39197..39292))
    /locus_tag="RP3-370M22.3-003"
    /note="match: cDNAs: AF062084"
    complement(join(4093..4154,7268..7510,32902..33450))
    /locus_tag="RP3-370M22.3-003"
    /standard_name="OTTHUMP0000028839"
    /codon_start=1
    /protein_id="CAI42712.1"
    /db_xref="GI:57209987"
    /db_xref="UniProt/TrEMBL:Q5H9F7"
    /translation="MAFRQVKNFVKXNSDIEIKVREATNSDPWGPSSSLMLDISDLT
    FNTSLISIMIMMLWRLNDHGKMMHWKYSKTLMDYLLIKNSKVTQHCRGFCNLQYQ
    LKDFDHIDGAGDQGYVIREKSKVITLMEPELCKREYACRROSTSLSPSKK
    OLGSNSLTLCTASATPDISASEKKYKLPKGRLLHNKTHGSRSDPNFVLLKKYIS
    <20799..21898
    /locus_tag="RP3-370M22.2-001"
    <20799..21898
    /locus_tag="RP3-370M22.2-001"
    <20799..21617
    /locus_tag="RP3-370M22.2-001"
    /standard_name="OTTHUMP0000028838"
    /codon_start=1
    /protein_id="CAI42711.1"
    /db_xref="GI:57209986"
    /db_xref="GOA:Q5H9F8"
    /db_xref="InterPro:IPR004192"
    /db_xref="InterPro:IPR005805"
    /db_xref="InterPro:IPR005806"
    /db_xref="InterPro:IPR006317"
    /db_xref="UniProt/TrEMBL:Q5H9F8"
    /translation="SVARSQFPALVPSATSRGVAGALRLPVOATVPATPEQPYLDLRL
    RPLRSRLDSQAVRRPLVASGVGLNPASVCSHTDYVDPFSERYRLRELVDSYSSS
    ESTARKSFSLVLTGVTGVAAYAAKNAVLSQIDRPOHDLDRVKKPEVILIGCTH
    KMAKKRKGKELPVHRHTQKEIKQBAVAELSQLDPOHDLDRVKKPEVILIGCTH
    GCVPLANNGDGGYVPCPGSHYDASGRIRLGPATLNLNEVTFEFTSDMWIVG"
    /locus_tag="RP3-370M22.1-001"
    join(46614..46832,92595..92686,101321..101412,
    105557..105676,111492..111660,113544..113774,
    114913..115035,116407..116844)
    /locus_tag="RP3-370M22.1-001"
    join(46614..46832,92595..92686,101321..101412,
    105557..105676,111492..111660,113544..113774,
    114913..115035,116407..116844)
    /locus_tag="RP3-370M22.1-001"
    join(46614..46832,92595..92686,101321..101412,
    105557..105676,111492..111660,113544..113774,
    114913..115035,116407..116844)
    /locus_tag="RP3-370M22.1-001"
    /standard_name="OTTHUMP0000028837"
    /codon_start=1

```

```
/protein_id="CA142713.1"
/db_xref="GI:57209988"
/db_xref="GOA:O75791"
/db_xref="InterPro:IPR000980"
/db_xref="InterPro:IPR001452"
/db_xref="UniProt/Swiss-Prot:O75791"
/translation="MEAVAKEDFTASGDELSPFTGVLKILSNQEMFNAELSGSQG
YVKNFIDIOFPKWEHGLSQAENLIMKEVGFFIIKASQSPGDFISVRHEDV
OHPKVMKNGKYNFLMTEKPPSLINKLVYPTNLSIQHOKI FLDRPREDOGHNSL
DRSQGPHLSGAVGERIRPSMNRKLSHPPTLPLQOHQPOPOVAPPOLOQOP
QQRYLQHHFHQERKGSLLINDHCGTGLGSENNALMERHRTDPLOLAAGVYRA
RALYDFALEDDLEGFHSGEVVLDSNPSMTWTRLNKLGLFPANVAVAMTR"
gene
/locus_tag="RP3-370M22.5-005"
/locus_id="110407"
/pseudo
109527..110407
/locus_tag="RP3-370M22.5-005"
/locus_id="110407"
/pseudo
/codon_start=1
join(140545..141033,283847,1:21027..31898)
/locus_tag="RP3-496C20.4-004"
join(140545..141033,283847,1:21027..21194,
283847,1:21746..21867,283847,1:23149..23822,
283847,1:31375..31898)
/locus_tag="RP3-496C20.4-004"
join(140545..141033,283847,1:21027..21194,
283847,1:21746..21867,283847,1:23149..23822,
283847,1:31375..31424)
/locus_tag="RP3-496C20.4-004"
/standard_name="OTTHUMP0000028840"
/codon_start=1
/protein_id="CA142715.1"
/db_xref="GI:57209989"
/db_xref="UniProt/Trembl:Q8NEG4"
/translation="MAESQLNCLDEAHVNEKTEQAAPFYCERRRALLEALGGEG
AYRLKEEQLRPLSPEKQALRAWSPYEDVAVPANAARQSKAKAAPAPVAPQ
ESLAWPDRSDTEVPLDGLWDTGTFYGVSVLTFTPHPKKAPHLKQVVRQMIQI
AOKVIAVMDLPTDGDIDODIADACRKPVYIILIDAGVYFLSMCDDLETPRI
RNIVRSVYGTGFWPMGRIGKTLSSRLMDGDVATAGSRFTMSHVDNLLILL
TGQVVEPDTRELYALSEEDLYRQSLAGRVLHSSYARLKLINPKYALVSGCR
HPPEMRMWAARQOEAGNPEQGEESGASARLSPFLDLVLTQVLPVPEPIR
LGEISQKGRVSHMRDLKPKSREAPSRNMGEARKEAARFSSRLRFRKRP
AAPNGMASSVSTETSEVFLTGKRPENSSADISGKTSPSSAKPSCVVIS"
ORIGIN
Query Match 25.5%; Score 510.4; DB 8; Length 143747;
Best Local Similarity 76.2%; Pred. No. 26-70;
Matches 808; Conservative 0; Mismatches 186; Indels 66; Gaps 12;
```

```
QY 1310 GCAGTTCTTAAA-----ACACAGTGAAGATTGCAATATACGTGACTCTTCTTCAT 1364
DB 3489 GCATTTTAAAGTAAAGACACATTAAGCTGCAACATCACTGCTCTTCTTAC 3548
QY 1365 GAAGATTTCTCTCTACTGATGATAGCTTTTGTATGACATTTTATGACAGTAACTTC 1424
DB 3549 AAAAGATTATCTGTACATGATGATGCTGTTTATACATTTTACCAAGTAACTTC 3608
QY 1425 TTGAAATTTGA-TCAATCTCTCAAAACCTGCTGCTTTAAACAATTAAATAT 1483
DB 3609 TTTCAAATTTGAGTCAATCTCTCTAATCTGCACTGGTTTATTACTAAGTTATGT 3668
QY 1484 AATATCTGAATTCATTTGTTGTCATTTTCAACAATTTTCAAGTCTTTCACAGAACTAG 1543
DB 3669 AATATCTGAATTTCTTTGTTGGCATTCACAAATGTTCAAGCTTCTTCA-CAGAGTATG 3727
QY 1544 ATTGCATCTCATTTCCGAGATGGAATCTTTGTCATCCATTAAGAAATTTCTCATCT 1603
DB 3728 ATTCAACTCAAGAACCATTT-----TCTTTGCTCATTCATTAAGAAATTTCTCATCT 3783
QY 1604 GTTCAAGTTTATCATGATGATGAGCAATPACAGTCAATGTTCTTCAGGCTCATCTTCACTT 1663
DB 3784 GTTCAAGTTTATCATGATGATGAGCAATPACAGTCAATGTTCTTCAGG-----CTTCACTT 3838
QY 1664 TTAATTCAGTCTCTCTGCTGTTTCTAACACATCTGCTGTTCTTCTTCCATTTGAAGTCT 1723
DB 3839 CTAATTCAGTCTCTCTGCTGTTTCTAACACATCTGCTGTTCTTCTTCCATTTGAAGTCT 3898
QY 1724 TGAACCTCTCCAACTCATCCATGAGGTTGGAATCGACTTCTTCAAAATTCCTGTATATA 1783
DB 3899 TGAACCTCTCCAACTCATCCATGAGGTTGGAATCGACTTCTTCAAACTCTGTATATG 3958
QY 1784 TTTATATTTTGS---ACCTCCATGATCATGATGATGTTCTTATGGAACCTGGAATGTGA 1840
DB 3959 TTGAAATTTTGGCTTCTTCCCATGATGATGATGATGATGATGATGATGATGATGATGATG 4018
QY 1841 ATCTTTTCAAAAGTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1900
DB 4019 ATCTTTTCAAAAGTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4073
QY 1901 TTCAATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1960
DB 4074 ATCTATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4133
QY 1961 AATTACTCTTGAATCATTTTCTGCAAAATGATGTTGTG 2000
DB 4134 AATCACTCTTGAATCATTTGCTGCAAAATGATGTTGTG 4173
RESULT 32
AC022101/c 173691 bp DNA linear PRI 29-NOV-2001
LOCUS AC022101 Homo sapiens chromosome 5 clone CTC-21005, complete sequence.
DEFINITION AC022101
VERSION AC022101.4 GI:17149373
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 173691)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 173691)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 173691)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
```





gene /db\_xref="taxon:9606"  
/chromosome="6"  
/map="q27"  
/clone="RP3-366N23"  
/clone\_1ib="RP3-366N23"  
/gene="UNC93A"  
/locus\_tag="RP3-366N23.1-001"  
join(29762..29925,32866..33047,34381..34610,36294..36419,42283..42482,44264..44399,46128..46259,53536..53602)  
/gene="UNC93A"  
/locus\_tag="RP3-366N23.1-001"  
/product="unc-93 homolog A (C. elegans)"  
/note="match: ESTs: T60316 T60341 W78931  
match: CDNAS: AJ508812"  
join(29839..29925,32866..33047,34381..34610,36294..36419,42283..42482,44264..44399,46128..46259,53536..53602)  
/gene="UNC93A"  
/locus\_tag="RP3-366N23.1-001"  
/standard\_name="OTTHUMP0000017623"  
/note="match: proteins: CAD48541"  
/codon\_start=1  
/product="unc-93 homolog A (C. elegans)"  
/protein\_id="CAI42706.1"  
/db\_xref="GI:57208158"  
/db\_xref="UniProt/TREMBL:Q5JZD6"  
/translation="MDSRLRVLVVSPFLLFTFAYGGLQSLQSSLYSEEGIGVTALS  
TLVGMILSMPLPLILBRLGKGTITLSMCGYVAFSGNPFAPWYTLIPSTILGL  
GAPPLMSQCTYLTITGNTTHAKKARKKQDNQYFGFLPILFSSGWMILSIVL  
GQPSQETLPEBOLTSQASDCMLATTTNSFORPSQCVTLITLGYAVLMAIFLO  
YRTVQRESSEKSKVPFWSLTLSFKLYRDRCLLILPLYSGLQGFSSSEYRS  
VLTCTIQFVGYVMICFSATDLCVLYGKVSQYTGRAVLVLAATHVSCMLALL  
WRPRADHLAVFVPSGLWADVADVMQTNALYGLPEKSKAEPANRYLMEALFVI  
AFGYSMPFVHVKLYTLGLVSLTWAVAGLVCEVSKNPPIRPAPGQVQADEDEIQT  
KM"  
63467..80236  
/gene="RP3-366N23.3"  
/locus\_tag="RP3-366N23.3-001"  
join(63467..63569,76996..77152,78454..80236)  
/gene="RP3-366N23.3"  
/locus\_tag="RP3-366N23.3-001"  
/note="match: ESTs: BG720252 B1662156 B1643431 B1560373  
match: CDNAS: AY026506"  
join(63523..63569,76996..77152,78454..80028)  
/gene="RP3-366N23.3"  
/locus\_tag="RP3-366N23.3-001"  
/standard\_name="OTTHUMP0000017623"  
/note="match: proteins: Q9BWV7"  
/codon\_start=1  
/protein\_id="CAD92564.1"  
/db\_xref="GI:31652294"  
/db\_xref="GeneW:21211"  
/db\_xref="GO:Q9BWV7"  
/translation="MRGRDLCSSTOSQALGRTTPATPLNIPSEANHTPEOPAGLG  
ALQAGVSIIPRRGRPTLEKKKKPHLMADEPSGALIKLIVREVETPAVVOV  
ILERGMNKPKQONABDWNLWMTSSPFMTESNYSKPMQNLNHSGLTKRKQCLA  
KILKMRMYGTSILYQFTPLTFVMPNDYKTPVAYEFOEOMQGTGASVWICDPAISL  
RGILIFSDPKDFPDMDYTVQKYSNPLIRPKCDLAIYCVTGFKXLTITTYVQEG  
LVRFATEKFDLSNLQNNVYALTNSSINKSGASYEKIKVEYIGGCWMTLSRFSYLSW  
DVDDLLMKKIHWMVILITLIALSVFPFANCFELFGPILIDIDNLKPLWLVNVSFA  
LTLDCSTLVLRKLVHDIILYLNGLENGREASNAATHGNSIDAAKSDRGGLDAP  
DLTPDLSFTSRMNEDESVVEKAVSVRPAAPASOLEGMSGODFHLSTREBOSK  
PCLBRHTPHKTLMPYASLFQSHSCCTKTSFVLSRGAAPDPQAGNPLVLPFMEAT  
LQASRNLVVKRIIDELQKLMKQHS"  
80171..80176  
/gene="RP3-366N23.3"  
/locus\_tag="RP3-366N23.3-001"  
80236  
/gene="RP3-366N23.3"  
/locus\_tag="RP3-366N23.3-001"  
85821..86402  
/locus\_tag="RP3-366N23.4-001"

CDS /pseudo  
85821..86402  
/locus\_tag="RP3-366N23.4-001"  
/note="match: proteins: P79935 Q13257"  
/pseudo  
/codon\_start=1  
complement(95380)  
join(complement(AL592444,18:16183..26342),  
complement(95382..97761))  
/gene="TCP10"  
/locus\_tag="RP3-366N23.5-001"  
join(complement(AL592444,18:2633..26342),  
complement(AL592444,18:24643..24787),  
complement(AL592444,18:22953..23168),  
complement(AL592444,18:19787..19924),  
complement(AL592444,18:18378..18366),  
complement(AL592444,18:17798..17909),  
complement(AL592444,18:16183..16283),  
complement(97644..97761),complement(95382..96645))  
/gene="TCP10"  
/locus\_tag="RP3-366N23.5-001"  
/product="c-complex 10 (mouse)"  
/note="match: ESTs: AA431681.1 B1828613.1  
match: CDNAS: U03399 U03399.1"  
complement(95407..95412)  
/gene="TCP10"  
/locus\_tag="RP3-366N23.5-001"  
join(complement(AL592444,18:22953..23168),  
complement(AL592444,18:19787..18366),  
complement(AL592444,18:18378..17909),  
complement(AL592444,18:16183..16283),  
complement(97654..97761))  
/gene="TCP10"  
/locus\_tag="RP3-366N23.5-001"  
/standard\_name="OTTHUMP0000017628"  
/note="match: proteins: Q9N019"  
/codon\_start=1  
/product="c-complex 10 (mouse)"  
/protein\_id="CAI42705.1"  
/db\_xref="GI:57208157"  
/db\_xref="UniProt/TREMBL:Q5JZD6"  
/translation="MEKTPAAEYVREDSNGEMPSLQOQITSLHOBIGROQSLWADI  
HRKLQSHDALTQKQRLREIRGLQROQWAGKPPASPIAGRSHTLALBPARKI  
SHLSADEDTLPKYAGRSQSAITLIGORSSSHLAPKPMXKTERINSKXTPPOEDRE  
KSPGRRODRSPAPGRTPGABRRGVSEDKIMPSRSRSGRRSPQASQATT  
LOBQTHAARGDRSSSVLGSSBGGFLSRVQDEPARSSAPDAERQKSTPAPAKKSCF  
PTGRWNISMRKRPRPYLMGQS"  
polyA\_signal  
CDS  
polyA\_signal  
ORIGIN  
Query Match 25.4% Score 507.6; DB 8; Length 98517;  
Best Local Similarity 75.6% Pred No. 6.2e-70;  
Matches 791; Conservative 0; Mismatches 204; Indels 51; Gaps 11;  
QY 977 TACAGCATACCTTGAGATACCTGTGGTGGTTCATACCAACAATAAACAATA 1036  
Db 92618 TACAGGACAGGCTTGAGATACCTGTGGTTCAGGACCAACAATA----- 92568  
QY 1037 TGCAGGAAGTGATATACAAATAAAGTAGTACCAAGCTCTTTGGCTTCCCAATGCAT 1096  
Db 92567 -----AAGCCCATATTCGACAGGACAGCAAGTACCAATATTTGGTCCCAAGTGCT 92514  
QY 1097 ATAAAGTTTGCTATATCTACACGATGTGCTGTTAAGTGCATAGTATGCTTAA 1156  
Db 92513 ATAAAGTTTGCTATATATATACGTAAGTGTATTAAGTGTATTAATAGCATTAATGCTTAA 92454  
QY 1157 AA-AAACACATACCTTAATTTTAAATGCTTATTT-ACATAAAATGCTAACATCAT 1213  
Db 92453 AATTAAGACATATCTTAATTTAAATAGTATGATTAAATAATGCTAATGATCATC 92394  
QY 1214 TGAAGATTGAGTGTGATCTTTTGGCTGGTGAAGGCTTTTCTTATGATGACG 1273  
Db 1273 TGAAGATTGAGTGTGATCTTTTGGCTGGTGAAGGCTTTTCTTATGATGACG 1273

Db 92393 TGAGCCCTCAGTGAATTAATGTTTGTCTGCTGGAGGAGGCTTGCTGGATGATGATG 92334  
 1274 AT-----CGGGGCTCAGGCTGCTAAGCTTGGGCTGGAGGTTCT----- 1318  
 Db 92333 CTGCTGACTGATCAGGGGTGAGGTTGCTGAAGGTGAGGTGAGGCAATATTCGAAA 92274  
 1319 TAAACACAGTGAAGATGCAATATCAGTTGACTCTCTCTTCATGAAAGATTTCTCTC 1378  
 Db 92273 TAAAGACCAATGAATATTACTGATGACTGACTCTCTCTCTCAAGAAAGTTTCTCTG 92214  
 1379 TAGGTGTGATGCTTTTGTATGATGATTTATGACAGTGAAGAACTTTTGAAGATGGA- 1437  
 Db 92213 TAGCATGCGATGCTGTTTATGATGATTTTACCACAGTACAACTTCTTTTAAATGAG 92154  
 1438 TCAATCCTCTCAACCCCTGCTC-TGCTTTAACACACTAAGTTATATATTTTGAATC 1496  
 Db 92153 TCAATCCTCTCAACCCCTGCTC-TGCTTTAACACACTAAGTTATATATTTTGAATC 92094  
 1497 CATTGTTGCTTCAACAAATTTTTCACAGTGTCTTACACAGAGTATGATTCATTT 1556  
 Db 92093 ATTTGTTGCTATTCATTAACGTTTACAGCATCTTACACAGAGTCAATTTCTCAAG 92034  
 1557 TCTCTGATGGAATCTTGTCTCATCCATTAAGAAAGAAATTCCTCATCTGTTCAAGTTTAT 1616  
 Db 92033 AAACCACTT-----TCTTCTCAACCATTAAGAAAGCACTCTGATCTGTTCAATTTTAT 91978  
 1617 CATTGATTCGACCAATACAGTCATGTTCTTCAAGGCTCTCATCTTTAATTTCAAGTTC 1676  
 Db 91977 CCTGAGATTGACGACATTCAGTCACATTTTGTAGGCCCACTTC-----TAAATCTAGTTC 91923  
 1677 TCTGTGCTGTTTACACATCTGTGCTCTCTCTCATTTGAAGTGTGAACCTCTCCAA 1736  
 Db 91922 TTTTGTATTTTCACTCTCATACAGTATCTTCTGACAGTGAAGGTTTAAACCCCTCAAA 91863  
 1737 GTCATTCATGAGGGTGTGATGATGACTTCTTCAAAATTCCTGTTAATTTATTT-ATTG 1794  
 Db 91863 ATCATTCATGAGGAATGAATCACTCTCTTAAACTCTGTTAATGTTGATTTGTTTC 91803  
 1795 ACCTCCCATGATCATGATGATTTCTTAAATGCACTGTGAATGCTTCTTCCAAAG 1854  
 Db 91802 TCCCTCCGATGATCATGATGATTTTAAATGCACTGTGAATGCTTCTTCTAGATG 91743  
 1855 GTTTCAATTTACTAGTCCAGATCATCCATCCAGAGATCCACTTTGAAGCCAGTTA 1914  
 Db 91742 GTTTTAAATGACTTTGCCAATTCAT-----CTGAGGAATCATCATCTTAAAGCCAGCTTA 91688  
 1915 TAGCCTTATGATGATGATTTCTTCAATATAAGGCTTGAAGTTGAATTTACTCTTGTAT 1974  
 Db 91687 TAGCCTTATGATGATGATTTCTTCAATATAAGGCTTGAAGTTGAATTTACTCTTGTAT 91628  
 1975 CCATTTTCTGCAAAATAGATGTTGTTG 2000  
 Db 91627 CCATGGGCTGCAAGATGATGTTG 91602  
 RESULT 34  
 AC084703/c 110907 bp DNA linear PRI 28-MAR-2002  
 LOCUS Homo sapiens chromosome 15, clone RP11-1152D21, complete sequence.  
 DEFINITION AC084703  
 AC084703 GI:19774531  
 VERSION HTG.  
 KEYWORDS Homo sapiens (human)  
 ORGANISM Homo sapiens  
 SOURCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Euthera; Euarchoptogilres; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 110907)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 15, clone RP11-1152D21  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 110907)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,  
 Bouhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
 Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
 Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karates,A., LaRoque,K.,  
 Lamazares,R., Lander,T., Lehoczy,J., Levine,R., Liu,C., Liu,G.,  
 Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McKernan,K.,  
 McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,  
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
 O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,  
 Pierre,N., Pisanzi,C., Pollara,V., Raymond,C., Riebeck,M., Riley,R.,  
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
 Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Straus,N., Subramanian,A., Talamas,U., Testaye,S., Theodore,J.,  
 Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,  
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
 Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (08-NOV-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 110907)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
 Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamei,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,  
 Lander,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicoll,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Rella,R., Riebeck,M., Riley,R., Riese,C., Rogov,P.,  
 Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 110907)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
 Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamei,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,  
 Lander,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicoll,R.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Rella,R., Riebeck,M., Riley,R., Riese,C., Rogov,P.,  
 Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 28, 2002 this sequence version replaced gi:18698799.  
All repeats were identified using RepeatMasker:  
Smith, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L10865  
Center clone name: L152\_D\_21  
-----

Only the first 110.9 kb of this clone are being submitted.  
The remainder overlaps accession number AC090868 [WICR project  
L13154].

## FEATURES

## source

1..110907  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone="RP11-1152D21"  
/clone\_11b="RP11-11 Human Male BAC"  
143..166  
/rpt\_family="AT\_rich"  
complement(167..489)  
/rpt\_family="MER2"  
490..512  
/rpt\_family="AT\_rich"  
complement(766..985)  
/rpt\_family="L2"  
1051..7210  
/rpt\_family="L1PA3"  
7986..8009  
/rpt\_family="AT\_rich"  
complement(8312..8382)  
/rpt\_family="MER105"  
complement(8383..8530)  
/rpt\_family="MER5A"  
complement(8531..8576)  
/rpt\_family="MER105"  
8632..8660  
/rpt\_family="AT\_rich"  
9055..9211  
/rpt\_family="FAM"  
9921..9956  
/rpt\_family="(A)n"  
complement(10326..10612)  
/rpt\_family="L1PA7"  
complement(10704..10749)  
/rpt\_family="MIR"  
complement(10818..10922)  
/rpt\_family="FLAM\_A"  
11092..11128  
/rpt\_family="AT\_rich"  
11315..11565  
/rpt\_family="Charlie7"  
12263..12512  
/rpt\_family="L1MA9"  
12824..12979  
/rpt\_family="L1M4C"  
complement(13134..13244)  
/rpt\_family="L1MB"  
13271..13680  
/rpt\_family="MLT11"  
13782..13804  
/rpt\_family="AT\_rich"  
complement(13824..13955)  
/rpt\_family="MIR3"

repeat\_region 13986..14025  
/rpt\_family="AT\_rich"  
repeat\_region complement(14176..14470)  
/rpt\_family="AluSc"  
repeat\_region 15104..15130  
/rpt\_family="AT\_rich"  
repeat\_region complement(15557..15615)  
/rpt\_family="MIR"  
repeat\_region complement(15651..15993)  
/rpt\_family="MLT11"  
repeat\_region complement(16131..16421)  
/rpt\_family="AluJb"  
repeat\_region complement(16486..16993)  
/rpt\_family="L3"  
repeat\_region complement(17003..17213)  
/rpt\_family="MER58C"  
repeat\_region complement(17214..17676)  
/rpt\_family="r3"  
repeat\_region complement(17677..17985)  
/rpt\_family="AluSc"  
repeat\_region complement(17986..18090)  
/rpt\_family="L3"  
repeat\_region 18091..23830  
/rpt\_family="L1PA7"  
repeat\_region 23834..23856  
/rpt\_family="TAAA)n"  
repeat\_region 23975..24193  
/rpt\_family="MER58A"  
repeat\_region 24352..24675  
/rpt\_family="L2"  
repeat\_region complement(25054..25564)  
/rpt\_family="MER119"  
repeat\_region complement(25710..25886)  
/rpt\_family="MER58A"  
repeat\_region 26048..26488  
/rpt\_family="L1MC4"  
repeat\_region 27129..27235  
/rpt\_family="MER103"

Query Match 25.3% Score 506.2; DB 8; Length 110907;

Best Local Similarity 75.3% Pred No. 9.7e-70; Matches 793; Conservative 0; Mismatches 208; Indels 52; Gaps 11;

QY 974 TAGTACAGGACATCTGAGAGTACTGTGGGTTGGTTCATACCAACCAATTAATACAA 1033  
DB 62627 TAACTCAGGTAACTTGAAGTACTGTGGGTTCAAGACCAACCAATTAATACAA 62568  
QY 1034 AATGCAAGAGTGCATATCACAAATGAAGTACACACAGTCTTTGGCTTCCAGTG 1093  
DB 62567 ATGTA-----GGAAATAAATGTAGGAATGATGATGAAGTTTAAAGTTTCCAGTA 62516  
QY 1094 CATATAAAGTTTGGCTTATATACACTGATGCTGTTAAGTGTCAATAGTTATGTC 1153  
DB 62515 CATATAAAGTATGTTTACATATACATGATGCTGTTAATATGCTATAGCATTTATGTC 62456  
QY 1154 TAAAAAAC-----ACATACCTTAATTTTAAATGCTTTATTAATAAAATGCTAACAA 1208  
DB 62455 TAAAAAACATATGATATACATCTTAATTTCAAAATCTTATGTAATAAAAATGCTAATGA 62396  
QY 1209 TCAATTTGACATTCAGTGAAGTTTAAATCTTTTGGCTGGTGAAG-----GT 1254  
DB 62395 TCACTGATCTTTCAGCAAGTGTATCTTTTGGCTGGTGAAGGGGCTTGGCCGTTA 62336  
QY 1255 CTTTTCCTATTGATGATGATGAGGGGGGTGAGG-TGCTGAAGCTTGAAGGCTGTGGCAG 1313  
DB 62335 TGTAAATGGCTCTGACATATCAAGGTATGGTGTCTGAAGGTGGATGGCTGTGGCGT 62276  
QY 1314 TTTCTTAAACAACAGTGAAGATTTGCAATATGATGATCTTCTTTCATGAAGATTT 1373  
DB 62275 AAAATCGACACAAAATGAATTTGCTCA-----TTGACCTTCTTTCATGAAGATTT 62220  
QY 1374 CTCTAGTGTGATGCTTTTATGATGATTTTATGACAGTGAACCTTCTTGAAT 1433

```

Db      62219 CTCGTAGCATGATGATCTGTTTGATGACATTTTACTTACAGTAAAGACTTCTTTCAGAAAT 62160
Qy      1434 TG-GATCAATCCCTGCAAAACCCCTGCTGCTTAAACAACCTAGTTAATATATATCTG 1492
Db      62159 TGAATTCCTGCTCTTCAAAACCCGCGGCTTATCACTAAGTTAAAGTATATCTCA 62100
Qy      1493 AATCCATGTTGTCATTTTCAAACTTTTCAAG--TGTCTTCAACGAGAGTATATCCAT 1550
Db      62099 AGTGTGTTGTGTCATTTTCAAGCAATTTTCAATCATCTTTTCAACGAGAGCTTTCCAT 62040
Qy      1551 CTCATTTCTGAGATGGAATCTTTGCTCATTCATTAAGAAGAAATTCCTCATCTGTTCAAG 1610
Db      62039 CTTAAGAAATATTT-----TCTTCTTCACTCCCTAAGAAAGTACCTCTCATTTATCAAG 61984
Qy      1611 TTTTATCATGATGATGACAGAAATAGCATGTCCTCAGGCCCATCTTCACTTTTAATTC 1670
Db      61983 TTGGTCAATGATGATGACAGCAATTCATGTCATCTTTAGGC-----TCCACTTTTAGTTC 61929
Qy      1671 CAGTCTCTTGTGCTGTTTCAACCATCTGTGTTCCCTCCCTCCATGGAAGCTTTGAACCT 1730
Db      61928 TGGTCTTTTGGCCCTTTCAACCATCTGCAAGTTCCTTCTTCACTGAAAGCTTTGAACCC 61869
Qy      1731 CTCGAAGTATCATGATGAGGTTGGAATCGAATCTTCTTCAAAATTCCTGTTAATATTAAT 1790
Db      61868 CTCGAAGTATCATGATGAGGTTGGAATCGAATCTTCTTCAAAATTCCTGTTAATATTAAT 61809
Qy      1791 TTGGA---CCTCCATGATCATGATGATGTTCTTAATGGAACCTGGAATGTAATCTTT 1847
Db      61808 TTGAGACTCTCCCAAGAAATCAAGAAATGTTCTTAATGATGATGATGATGATCTT 61749
Qy      1848 CCAAAAGTTTTCATTTTACTTACTAGTCAGATCCATCCATCCAGAGATCCACTTTCAGT 1907
Db      61748 CTAGAAAGTTTTCAGTCTACTTCTTCCCAAGTCA-----GTAGAGAAATCATATTTTAAG 61694
Qy      1908 CCAAGTATAGCCTTATGATGATGATGATGTTCTTCAATTAATAAGCTTGAAGTTGAATTAAT 1967
Db      61693 GGAGCTAAAGCCTTATGAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 61634
Qy      1968 CCTGATTCATTTTCTGCAAAATAGATGTTG 2000
Db      61633 CCTGCTCATGAGACGAGAAATGATGTTG 61601

RESULT 35
AC013752/c 183794 bp DNA linear PRI 29-DEC-2002
LOCUS      Homo sapiens chromosome 15, clone RP11-396J17, complete sequence.
DEFINITION AC013752
ACCESSION  AC013752.10 GI:27413935
VERSION     HTG.
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 183794)
Hom sapiens chromosome 15, clone RP11-396J17
Unpublished
2 (bases 1 to 183794)

```

```

TITLE      Direct Submission
JOURNAL    Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
REFERENCE  Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS    Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Allen, N.,
            Anderson, S., Barna, N., Baskien, V., Boguslavsky, L., Boukhalter, B.,
            Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
            Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
            Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
            Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
            Ginos, B., Goyette, M., Graham, L., Grand-Pierre, N.,
            Hago, B., Hefford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
            Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
            Lamas, R., Lander, E., Lehotzky, J., Levine, R., Liu, G.,
            Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
            McCarty, M., McEwan, P., McKernan, K., McPeeters, R., Meldrum, J.,
            Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
            Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
            Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
            Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,
            Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
            Seaman, S., Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N.,
            Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testfaye, S.,
            Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J.,
            Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
            Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE      Direct Submission
JOURNAL    Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome
REFERENCE  Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS    Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
            Barna, N., Baskien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
            Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
            Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
            Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
            Gardyna, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B.,
            Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
            Kamat, A., Karatas, A., Kells, C., Lander, E., Levine, R.,
            Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
            Matthews, C., McCarty, M., Meldrum, J., Meneus, L., Mihova, T.,
            Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicoll, D., Oliver, J.,
            Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
            Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
            Riese, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R.,
            Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
            Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K.,
            Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
            Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE      Direct Submission
JOURNAL    Submitted (29-DEC-2002) Whitehead Institute/MIT Center for Genome
REFERENCE  Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Dec 29, 2002 this sequence version replaced gi:21622793.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            Project Information
            Center project name: 12432
            Center clone name: 396_J_17

            ----- Location/Qualifiers
            1. 183794
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="15"
            /map="15"
            /clone="RP11-396J17"

```

```
/clone_lib="RPCT-11 Human Male BAC"
repeat_region      48..160
/rpt_family="MIR"
repeat_region      complement(304..572)
/rpt_family="L1R16C"
repeat_region      1091..1186
/rpt_family="MIR"
repeat_region      complement(1197..1495)
/rpt_family="AluJo"
repeat_region      1497..1651
/rpt_family="MIR"
repeat_region      complement(1653..1838)
/rpt_family="L2"
repeat_region      1839..2257
/rpt_family="MSTB"
repeat_region      complement(2258..2315)
/rpt_family="L2"
repeat_region      complement(2397..2698)
/rpt_family="L2"
repeat_region      complement(3400..3588)
/rpt_family="L1PA16"
repeat_region      complement(3590..5413)
/rpt_family="L1PA16"
repeat_region      5414..5517
/rpt_family="TTTC)n"
repeat_region      complement(5518..5804)
/rpt_family="L1PA16"
repeat_region      complement(7391..7969)
/rpt_family="MER68"
repeat_region      8172..8322
/rpt_family="PRAM"
repeat_region      8326..8378
/rpt_family="TAAA)n"
repeat_region      9897..10803
/rpt_family="L2"
repeat_region      11116..11227
/rpt_family="MIR"
repeat_region      11419..11716
/rpt_family="AluJo"
repeat_region      12056..12212
/rpt_family="PRAM"
repeat_region      12365..12506
/rpt_family="L2"
repeat_region      complement(13509..13604)
/rpt_family="MIR3"
repeat_region      14272..14440
/rpt_family="MER117"
repeat_region      complement(15242..15443)
/rpt_family="L1MEC"
repeat_region      15698..15875
/rpt_family="L1R47A"
repeat_region      complement(15878..16264)
/rpt_family="MSTB"
repeat_region      complement(16265..16482)
/rpt_family="L1MEC"
repeat_region      complement(16483..16787)
/rpt_family="AluJo"
repeat_region      complement(16788..17182)
/rpt_family="L1MEC"
repeat_region      complement(17237..17283)
/rpt_family="L1MEC"
repeat_region      complement(17284..17593)
/rpt_family="AluJo"
repeat_region      17601..17626
/rpt_family="(TG)n"
repeat_region      complement(17627..18455)
/rpt_family="L1MEC"
repeat_region      19065..19124
/rpt_family="(TATATG)n"
repeat_region      complement(19217..19428)
/rpt_family="MER106A"
repeat_region      19724..19779
/rpt_family="AT_rich"
```

```
repeat_region      19918..20460
/rpt_family="L1R9"
repeat_region      complement(20465..20783)
/rpt_family="L1M4"
repeat_region      complement(20784..21266)
/rpt_family="L1R40a"
repeat_region      complement(21267..21353)
/rpt_family="L1M4"
repeat_region      21647..21696
/rpt_family="(CA)n"
repeat_region      complement(21949..22126)
/rpt_family="MER5B"
repeat_region      23039..23392
/rpt_family="MT1A1"
repeat_region      complement(23809..24025)
/rpt_family="L1MC4"
repeat_region      complement(24161..24519)
/rpt_family="THE1C"
repeat_region      complement(26065..26217)
/rpt_family="MER5A"
repeat_region      complement(26239..26334)
/rpt_family="L1M5"
repeat_region      26705..26733
/rpt_family="AT_rich"
repeat_region      26775..26823
/rpt_family="MER111"

Query Match      25.3%; Score 506.2; DB 8; Length 183794;
Best Local Similarity 75.3%; Pred. No. 8,1e-70;
Matches 793; Conservative 0; Mismatches 208; Indels 52; Gaps 11;

QY      974 TAGTACAGGATACCTTGAGAGTACTGTGGGTTTGTTCCATACCAACCAATTAATACAA 1033
DB      163619 TAATGACAGGTATACCTTAGAGTACTGTGGGTTTCAAGACCAACATTAAGCAA 163560
QY      1034 AATGCAAGAGTGATATCACATTAAGTAGTCACACAAAGCTTTTGCTGCCAGTG 1093
DB      163559 ATGTG-----GGAAATAAGTAGAGATGATGATGAAGTTTATGTTCCAGTA 163508
QY      1094 CATATTAAGAGTTTGCTTATATACACAGTAGTCTGTTAAGTGCAATAGTATATGTC 1153
DB      163507 CATATTAAGAGTTTATGTTTACATATACGTATGATCATATTAAGTGCTATATATGTC 163448
QY      1154 TAAAAAACC-----ACATACCTTAATTTTAAAGCTTTATTAATCTAATAAATGCTAACAA 1208
DB      163447 TAAAAAACCATATGATCATATCTTAATTCAAAATATCTTTAGTAATAAATGCTAATGA 163388
QY      1209 TCATTTGACATTCAGTGAGTGTATCTTTTGGCTGGTGAAG-----CT 1254
DB      163387 TCATCTGATCTTTCAGCAAGTGTATCTTTTGGCTGGAGGAGGTCTTGCCGTGA 163328
QY      1255 CTTTCTTAATGATGATGATCGGGGGGTGACAG--TGCTGAAGCTTAGGGTGGCTGGACAG 1313
DB      163327 TGTATATGCTCTCTGACAGATTAAGGTATGTTGCTGAAGTTGGATGGCTGGCGCT 163268
QY      1314 TTTCTTAATAACAACAGTAGTAAGTGCATATATGATGATGATCTCTTCATGAAAGATT 1373
DB      163267 AAAACGACAGCAAAATGAAATTTGCTCA-----TTGACCTCTTTTATGAAAGATT 163212
QY      1374 CTCCTAGTGTGATGCTTTTGTATAGATTTTATGACAGTGAAGCTTTTGAAAT 1433
DB      163211 CTCCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 163152
QY      1434 TG-GATGATTCCTGCAACCGTGCCTTAAACAACCTAAGTTAATATATATTCG 1492
DB      163151 TGAATTCGCTCTCTCAAAACCGTGCCTTAAACAACCTAAGTTAATATATTCG 163092
QY      1493 AATGCAATGTTGATTCATTAACAATTTTACAG--TGCTGACGAGAGTATTCAT 1550
DB      163091 AATTGTTTGTGATTCATTAAGCAATTTTATGATGATCTTTTACCAAGAGAGTTCAT 163032
QY      1551 CTCATTTCTGAGATGGAATCTTTGCTATTCATTAAGAAATTCATCTGTTCAAG 1610
DB      1551 CTCATTTCTGAGATGGAATCTTTGCTATTCATTAAGAAATTCATCTGTTCAAG 1610
```

Db 163031 CTTAGAACTATT-----TCTTCTCATCCGTAAGATGACCTCTCATCTATTCAAG 162976  
QY 1611 TTTATCATGAGATGAGCAATATACATGCTTCAAGGCTCAGCTTCACTTTAATTC 1670  
Db 162975 TTTGTCATGAGATGAGCAATATACATGCTTCAAGGCTCAGCTTCACTTTAATTC 162921  
QY 1671 CAGTCTCTGCTGCTTCTTCAACACATCTGTGCTTCTCTCATTTGAAGCTTGAAGCT 1730  
Db 162920 TGTCTTTTGGCTCTTCTTCAACACATCTGTGCTTCTCTCATTTGAAGCTTGAAGCT 162861  
QY 1731 CTCGAAGTCAATGAGGCTTGAATTCGACTTCTTCAAAATCTCTTTAATTTAAT 1790  
Db 162860 CTCGAAGTCAATGAGGCTTGAATTCGACTTCTTCAAAATCTCTTTAATTTAAT 162801  
QY 1791 TTTGA---CCTCCATGAAATCATGAAATGCTTCAATGAGGCTTGAATGCTTCTT 1847  
Db 162800 TTTGACCTCTCTCCATGAAATCATGAAATGCTTCAATGAGGCTTGAATGCTTCTT 162741  
QY 1848 CCAAAAGTTTCAATTTACTTATGTCAGATTCATCCATCCAGAGATCCACTTTCAATG 1907  
Db 162740 CTAGAAAGTTTCAAGTCTACTTCTTCCAGATCCA-----GTAGAGGATCATCTATTAG 162686  
QY 1908 CCAAGTATAGCTTATGAAATGATTTCTTCAATTAATAGGCTTGAAGTTGAATTAAT 1967  
Db 162685 GCAGCTAAAGCTTATGAAATGATTTCTTCAATTAATAGGCTTGAAGTTGAATTAAT 162626  
QY 1968 CCTGATCATTTCTGCAAAATAGATGTTGTC 2000  
Db 162625 CCTGCTCATGAGCTGCAGAAATGATGTTGTC 162593

RESULT 36  
AC022028/16320 bp DNA linear PRI 22-MAY-2002  
LOCUS AC022028 Homo sapiens chromosome 10 clone RP11-93H12, complete sequence.  
DEFINITION AC022028  
AC022028 GI:21070552  
VERSION AC022028.7  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 156320)  
REFERENCE  
AUTHORS Smith,D.R.  
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome  
JOURNAL Sequence Data  
REFERENCE  
AUTHORS Smith,D.R.  
TITLE Direct Submision  
JOURNAL Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
3 (bases 1 to 156320)  
REFERENCE  
AUTHORS Smith,D.R.  
TITLE Direct Submision  
JOURNAL Submitted (18-JAN-2002) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
4 (bases 1 to 156320)  
REFERENCE  
AUTHORS Smith,D.R.  
TITLE Direct Submision  
JOURNAL Submitted (22-MAY-2002) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
On May 22, 2002 this sequence version replaced gi:18201761.  
FEATURES  
source  
1..156320  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-93H12"  
/clone\_1ib="RP11-93H12"  
ORIGIN

Query Match 25.3%; Score 505.2; DB 8; Length 156320;  
Best Local Similarity 75.5%; Pred. No. 1.2e-69;  
Matches 743; Conservative 0; Mismatches 208; Indels 33; Gaps 8;  
QY 1034 ATATGCAAGATGATATACATTAATTAAGTGAACACAGATCTTTGGCTCCAGTG 1093  
Db 112548 ATCTCAATTAAGTGAATATGCAATTAATTAAGTGCAGAAATTTTGGTTTCCAGTG 112489  
QY 1094 CATATTAAGTTTGGCTTATCTACATCTAGTCTGTTAAGTGTCAATAGTTATGTC 1153  
Db 112488 CATATTAAGTTTGGCTTATCTACATCTAGTCTGTTAAGTGTCAATAGTTATGTC 112429  
QY 1154 TAAAAAACAATACATCTTAAATTTAAATGCTTTTATCTTAAATAATGTAACATCAT 1213  
Db 112428 TAAAAAATATATATCTTAAATTTAAATGCTTTATCTTAAATAATGTAACATCAT 112369  
QY 1214 TGAGCATTCAGTATGATTTATCTTTTGGCTGTGAAGTCT-----TTTCTTAT 1264  
Db 112368 TGAGCATTCAGTATGATTTATCTTTTGGCTGTGAAGTCTTTGCTCAAAAGTTGATGTC 112309  
QY 1265 TGATGATGATGGGGGTGAGTCTGAAGCTTTAGGCTGTGTGCACTTTCTTAA-- 1322  
Db 112308 TCTGATGATGAGTGTGATTTGCTTAAAGTGAAGTGTGTGCACTTTCTTAAAT 112249  
QY 1323 ---ACAAAGTGAAGTATCAATCAATCAATGATCTCTCTTCAATGAAGATTTCTCT 1379  
Db 112248 GAGCAACAGTAACTTGCACATCAATGATCTCTTCTTCAATGAAGATTTCTCTCT 112189  
QY 1380 AGTGTGATGCTTTTGTATGATCTTTTATGACATTTATGACAGATGAACCTTTGAAATTTGA-T 1438  
Db 112188 AGCATGTATGATTTGTTTGTATCTTATTTACTACAGTGAACCTTTGTAAGCTTGGAGT 112129  
QY 1439 CAATCTCTCAAACTCTGCTCTGCTTTTAAACAACCTTAAGTAAATATCTGAATCA 1498  
Db 112128 CACACTCTCAAACTCTGCTCTGCTTTTAAACAACCTTAAGTAAATATCTGAATCA 112069  
QY 1499 TTGTTGTATTTCAACAATTTTCAAGTCTTTCAACAGATTAATTCATCTCATTTTC 1558  
Db 112068 TTGTTGTATTTCAACAATTTTCAAGTCTTTCAACAGATTAATTCATCTCATTTTC 112009  
QY 1559 CTGAGATGAATCTTTGCTCATCTCAATGAAGAAATTCCTCATCTGTTCAAGTTTATCA 1618  
Db 112008 GCCACTT-----TCTTGTCTCATCTGTGAAGAAATTTTCAACATTTAAGTTGATCA 111953  
QY 1619 TGAGATGCAAGATTAAGTCTTCAAGGCTCACTCACTTTAATTTCAAGTTCTC 1678  
Db 111952 TGAGATGCAAGATTAAGTCTTCAAGGCTCACTCACTTTAATTTCAAGTTCTC 111898  
QY 1679 TTGCTGTTTCAACACATCTGTGCTTCTTCTCAATGAAATCTTGAACCTCTCAAGT 1738  
Db 111897 TTGCTGTTTCAACACATCTGTGCTTCTTCTCAATGAAATCTTGAACCTCTCAAGT 111838  
QY 1739 CATCCATGAGGCTTGAATTCGACTTCTTCAAAATCTCTGTTAATTTATTTTGAAGCT 1798  
Db 111837 CATCCATGAGGCTTGAATTCGACTTCTTCAAAATCTCTGTTAATTTATTTTGAAGCT 111778  
QY 1799 CC---CATGAATCATGAATGTTCTTAATGGAACCTGGAATGTTGAATCTTTTCAAAAG 1855  
Db 111777 CCTCTCAAGATTAAGATTTCTGAATGGAACCTGGAATGTTGAATCTTTTCAAAAG 111718  
QY 1856 TTTTCAATTTTACTTGTGCAATTCATTCATCAAGAGATCCACTTTCAATGCCAGTTAT 1915  
Db 111717 -TTTCAATTTTACTTGTGCAATTCATTCATCAAGAGATCCACTTTCAATGCCAGTTAT 111664  
QY 1916 AGCTTATGAAATGATTTCTCAATTAATTAAGGCTTGAAGTGAATTAATCTCTGATC 1975  
Db 111663 AGCTTATGAAATGATTTCTCAATTAATTAAGGCTTGAAGTGAATTAATCTCTGATC 111604  
QY 1976 CATTTTCTGCAAAATAGATTTGT 1999  
Db 111603 CATGGCTGCAGAAATGATTTGT 111580

```

RESULT 37
LOCUS      AL139142      159771 bp      DNA      linear      PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP3-436N22 on chromosome 1q24.1-25.3
              Contains the PRDX6 gene for peroxiredoxin 6, gene MG43026, a novel
              gene, the 3' end of a novel gene (FLJ45235) and two Cpg islands,
              complete sequence.
ACCESSION  AL139142      GI:15620584
VERSION    AL139142.10
KEYWORDS   HTG; FLJ45235; MGCA3026; PRDX6.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
            1 (bases 1 to 159771)
REFERENCE  Tracey,A.
            Direct Submission
            Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
            Clone requestes: clonerequest@sanger.ac.uk
            On Sep 14, 2001 this sequence version replaced gi:15590740.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WormPep; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            This sequence
            was generated from part of bacterial clone configs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP3-436N22 is from the library RPc1-3 constructed by the group of
            Pieper de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pcYPAC2
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vegas@sanger.ac.uk
            -----
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest
            except on the rare occasion of the clone being a YAC.
FEATURES             location/Qualifiers
     source           1. 159771
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCLP704N2436"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q24.1-25.3"
                     /clone="RP3-436N22"
                     /clone_11b="RPc1-3"
     misc_feature     1
                     /note="Clone left end: RP3-436N22"
                     join(3133..3359,7193..7349,11228..11374,12122..12268,
                     13602..14674)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-001"
                     join(3133..3359,7193..7349,11228..11374,12122..12268,
                     13602..14674)
                     /gene="PRDX6"
                     /locus_tag="RP3-436N22.1-001"
                     /product="peroxiredoxin 6"
                     /note="match: CDNAs: Em:BC008738.1 Em:BC017522.1
                     Em:BC025421.2 Em:BC035857.1 Em:BC053550.1 Em:DL4662.1"
                     join(3197..3359,7193..7838)
gene

```

/gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-002"  
 join(3197..3359,7193..7838)  
 /gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-002"  
 /product="peroxiredoxin 6"  
 /note="match: ESTs: Em:CB988834.1"  
 join(3365..3359,7193..7349,11228..11374,12122..12268,  
 13602..13730)  
 /gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-001"  
 /standard\_name="OTTHUMP0000032693"  
 /note="match: proteins: Sw:O08709 Sw:O35244 Sw:O77834  
 Tr:AAP21829 Tr:Q7SYQ4 Tr:Q7SYT1"  
 /codon\_start=1  
 /product="peroxiredoxin 6"  
 /protein\_id="CAI20936.1"  
 /db\_xref="GI:56204402"  
 /translation="MGGILLGDPVAPNFEANTVGRIRPHDFLGGWGLIFSPHDFPT  
 PVCTTEIGRAAKAPAPFARKNNKTLALSLDSVEDHILANSGKIDINANCCEPTKLEPPI  
 IDDNRELRLIQLMDLPDAEKDEKMPETLRAVVFPGPDKKLSTLYPTTRRNDEI  
 LRVVISIQLAERKRAVPDMKGDSPMALPTLPEEARKLFPKGVFTKLEPSGKYL  
 RYTPQP"  
 join(7066..7349,11228..11374,12122..12268,13602..13882)  
 /gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-003"  
 join(7066..7349,11228..11374,12122..12268,13602..13882)  
 /gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-003"  
 /product="peroxiredoxin 6"  
 /note="match: ESTs: Em:AV751940.1 Em:BG025204.1"  
 13882  
 /gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-003"  
 14650..14655  
 /gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-001"  
 14674  
 /gene="PRDX6"  
 /locus\_tag="RP3-436N22.1-001"  
 complement(join(26331..26964,29133..29193,31114..31198,  
 32723..32908,35435..35566,43404..43571,47168..47266,  
 49836..49961,50646..50794,52530..52581,55772..55921,  
 59476..59664,60351..60549,61697..61816,62809..62906,  
 73556..73651,74256..74419,80600..80681,83207..83375,  
 99049..99192,102528..102627,107714..107875,109373..109489,  
 113532..113697,123771..123899,125984..126084,  
 127517..127722,128618..128961))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-001"  
 complement(join(26331..26964,29133..29193,31114..31198,  
 32723..32908,35435..35566,43404..43571,47168..47266,  
 49836..49961,50646..50794,52530..52581,55772..55921,  
 59476..59664,60351..60549,61697..61816,62809..62906,  
 73556..73651,74256..74419,80600..80681,83207..83375,  
 99049..99192,102528..102627,107714..107875,109373..109489,  
 113532..113697,123771..123899,125984..126084,  
 127517..127722,128618..128961))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-001"  
 /product="novel protein"  
 /note="match: CDNAs: Em:AK128104.1 Em:BC042592.1"  
 complement(26331)  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-001"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921,59476..59664,60351..60385,  
 61697..61816,62825..62906,73556..73651,74256..74419,  
 80600..80681,83207..83375,99049..99192,102528..102627,  
 107714..107785))  
 /gene="RP3-436N22.2"  
 /locus\_tag="RP3-436N22.2-002"  
 complement(join(26337..26964,29133..29203,31114..31175,  
 52938..53293,55772..55921



52938.. 53293,55772.. 55921,59476.. 59664,60351.. 60385,  
61697.. 61816,62825.. 62906,73556.. 73651,74256.. 74419,  
80600.. 80681,83207.. 83375,99049.. 99192,102528.. 102627,  
107714.. 107785))  
/gene="RP3-436N22.2"  
/locus\_tag="RP3-436N22.2-002"  
/locus\_tag="RP3-436N22.2-002"  
/product="novel protein"  
/note="match: cDNAs: Em:AB072793.1"  
complement(join(26337.. 26964,29133.. 29193,31114.. 31198,  
32723.. 32908,35435.. 35566,43404.. 43571,47168.. 47266,  
4836.. 49961,50646.. 50794,52530.. 52581,55772.. 55921,  
59476.. 59664,60351.. 60549,61697.. 61816,62809.. 62906,  
73556.. 73651,74256.. 74419,83207.. 83375,97369.. 97595,  
99049.. 99192,102528.. 102618))  
/gene="RP3-436N22.2"  
/locus\_tag="RP3-436N22.2-003"  
complement(join(26337.. 26964,29133.. 29193,31114.. 31198,  
32723.. 32908,35435.. 35566,43404.. 43571,47168.. 47266,  
4836.. 49961,50646.. 50794,52530.. 52581,55772.. 55921,  
59476.. 59664,60351.. 60549,61697.. 61816,62809.. 62906,  
73556.. 73651,74256.. 74419,83207.. 83375,97369.. 97595,  
99049.. 99192,102528.. 102618))  
/gene="RP3-436N22.2"  
/locus\_tag="RP3-436N22.2-003"  
/product="novel protein"  
/note="match: cDNAs: Em:BX648119.1"  
complement(26337.. 26357)  
/locus\_tag="RP3-436N22.2-003"  
complement(26352.. 26357)  
/gene="RP3-436N22.2"  
/locus\_tag="RP3-436N22.2-003"  
complement(join(26961.. 26964,43404.. 43571,47168.. 47266,  
4836.. 49961,50646.. 50794,52530.. 52581,55772.. 55921,  
59476.. 59664,60351.. 60549,61697.. 61816,62809.. 62906,  
73556.. 73651,74256.. 74419,80600.. 80681,83207.. 83375,  
99049.. 99192,102528.. 102627,107714.. 107875,109373.. 109489,  
113532.. 113697,123771.. 123899,125984.. 126084,  
127517.. 127643))  
/gene="RP3-436N22.2"  
/locus\_tag="RP3-436N22.2-001"  
/standard\_name="ORTHUMP00000032688"  
/note="match: proteins: Tr:Q86UR3 Tr:Q95LJ5"  
/codon\_start=1  
/product="novel protein"  
/protein\_id="CAI20938.1"  
/db\_xref="GI:56204404"  
/db\_xref="GOA:O5TAH2"  
/db\_xref="InterPro:IPR000595"  
/db\_xref="InterPro:IPR005820"  
/db\_xref="UniProt/TrEMBL:O5TAH2"  
translation="MSSYFMAQNESNRDPLCGPADYLVKEHFTLVICFIVLGL  
LMCKLNCSEVILITLISGVIGHNRAVHOIVYELKTSFSLYSYSPILIF  
WVALDFETLKKFMWOVLNGLISFSTSLIIIGVVFKENDSDLOSCLFSTLIG  
IIDPRVNSLKTIGISKYIDILIGBSLILCSIASIPFGNRGRHRSIRDLHVG  
IELSDYLDGSLIFGTWCATIDCILLADVFSNMLTNVILCFSNVMTFFIVEFLVMSGT

Query Match 25.3%; Score 505.2; DB 8; Length 159771;  
Best Local Similarity 74.3%; Pred. No. 1.2e-69;  
Matches 780; Conservative 0; Mismatches 238; Indels 32; Gaps 10;

QY 965 AGGCGTAGTAGTACAGGCATACCTTGGAGATAGTGGTGGTTGGTTCATACCAACACACA 1024  
DB 79697 AGAGATTTCATATACAGGCATACCTTGGAGATATTCAGATTCAATTCAGACCAACGTGA 79638  
QY 1025 ATATATCAATATATGCAAGAGATATCAATATTAAGAGATCAACAAAGCTTTTGGC 1084  
DB 79637 CAAGCAATATATGCAATA---ATATCAATATTAAGAGATCAATATTAATATATGTGT 79582  
QY 1085 TTCCAGTGCATATAAAGTTTGTCTTATACACTAGTGTAGTCTGTTAAGTGTGCA---A 1141  
DB 79581 TTCCAGTGCATATAAAGTTTGTCTTATACACTATACACTAGTGTATTAATATCAAAAGCA 79522

QY 1142 TAGGTATGTCTAAAAAACACATACCTTAATTTTAAAAATGCTTATTAAT-AAAAAT 1200  
DB 79521 CAGGTCTTAAGAAAACATATTTACATACCTTAATGAAGAAAATGCTTTCTGCTAAAAAAT 79462  
QY 1201 GCTAACATCATTTGAGCATTCAGTGTGTAATCTTTTTCGTGTGAGAGTCTTTTC 1260  
DB 79461 GCTAACATCATTCGTGAGCTTCAGCAATTCATTAATCTTTTTCGTGTGAGAGTGTG 79402  
QY 1261 TTATTTGATGACCTGA---TGGGGGGTCAAGTGTGAACTTAAAGTGGCTGTGCACTTC 1317  
DB 79401 CTGCTGACTGATCAATCAATCAGGTGTGTGTGTGTAAGATTTGGCAGCTGTGCAATTTCT 79342  
QY 1318 TTTAAAAACAACATGAAGATTTGCAATAGTGTGATGCTCTCTTCAATGAAGATTTCTCT 1377  
DB 79341 TTTAAGCAACATGAAGATTTGCAATAGTGTGATGCTCTCTTCAATGAAGATTTCTCT 79282  
QY 1378 CTAGTGTGTGATGCTTTTGTGATGATTTTATGACAGTGAAGATTTTGAAGATTTGAG- 1436  
DB 79281 GTAGCATGAGTGTGCTTTGATGATGATTTTATCATAGATGAGCTTTTCAAAATTTGA 79222  
QY 1437 ATCAATCTCTCAAAACC---TGTCTGTCTTAAACAACCTTAATATATATTTTGA 1493  
DB 79221 ATAAATCTCTCAAAACCCTGTGCACTTATGCACTTATGCACTTATATATATTTCTAA 79162  
QY 1494 ATTCATTTGTTCATTTTCAAAATTTTCAACATGCTTTCACAGAGATGATTTCACTTC 1553  
DB 79161 ATCTTTGTGTCTTCAACATGTTTCAACATGTTTCAACATGTTTCAACAGATGATTTCA 79102  
QY 1554 ATTTCTGATGATGATCTTTGCTGATTCATTAAGAAATTTCTCATCTGTTCAAGTTT 1613  
DB 79101 AAGAAAGCACTTCG---TTTACTATCTGTATGAAGAACTTCATCTCATGCAAGTTT 79046  
QY 1614 TATCATGATTTGACAGCAATATGATGCTTTCAAGGCTCACTTCACTTTAATTCAG 1673  
DB 79045 TATCTGATGATTTGACAGCAATATGATGCTTTCAAGGCTCACTTCACTTTAATTCAG 78991  
QY 1674 TTCTGTGTGTGTTTCAACATCTGTGTTCTTCTTCTTCACTTGAAGTCTTGAACCTTC 1733  
DB 78990 TTCTGTGTGTGTTTCAACATCTGTGTTCTTCTTCTTCACTTGAAGTCTTGAACCTTC 78931  
QY 1734 CAATCATCATGAGGTTTGAATGATGCTTTTCAAAATTTCTGTTAATTTAATTTT 1793  
DB 78930 AAGATCATCTAAGAAAGTTAATTAATTAATTAATTAATTAATTAATTAATTTT 78871  
QY 1794 GACCTC---CATGATCATGATGATTTCTTAATGACACCTGATGATGATTTTCA 1850  
DB 78870 GACCTCCTGATGATGATTTCTTAATGATGATGATGATGATGATGATGATTTTCA 78811  
QY 1851 AAGGTTTTCATTTTACTTACGTCAGATTCATTCATTCATTCATTCATTCATTCATTC 1910  
DB 78810 TGAGGTTTTTTTATTTATTTTCCCTACATTCAT---CAGAGAAATCATATCTATGACA 78756  
QY 1911 GTTATAGCTTATGATGATGATTTTCTCAATATATAGGCTTGAAGTGAATTTACTCT 1970  
DB 78755 GCTATAGCTTATGATGATGATTTTCTCAATATATAGGCTTGAAGTGAATTTACTCT 78696  
QY 1971 TGATTCATTTTCTGCAAAATAGATGTTGTG 2000  
DB 78695 TGATTCATTTTCTGCAAAATAGATGTTGTG 78666

RESULT 38  
AC021635  
LOCUS AC021635 162163 bp DNA linear HTG 24-AUG-2002  
DEFINITION Homo sapiens chromosome 3 clone RP11-418M15 map 3, WORKING DRAFT  
SEQUENCE, 9 unordered pieces.  
AC021635  
VERSION AC021635.3 GI:7230225  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 162163)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome 3, clone RP11-418M15  
Unpublished  
2 (bases 1 to 162163)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
Boguslavsky, L., Bouhagalter, B., Brown, A., Burkett, G., Castle, A.,  
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenebor, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howard, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,  
Lander, T., Lehoczy, J., Levine, R., Liu, C., Locke, K.,  
Macdonald, P., Marguis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 162163)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguslavsky, L., Bouhagalter, B., Brown, A., Burkett, G.,  
Campopiano, A., Castle, A., Choepe, Y., Colangelo, M., Collins, S.,  
Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,  
Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howard, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,  
Klein, J., Laroque, K., Lamazares, R., Lander, T., Lehoczy, J.,  
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marguis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,  
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Munga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 12, 2000 this sequence version replaced gi:6893683.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RX/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5968

Center clone name: 418\_M15

----- Summary Statistics

Sequencing vector: M13, M7815, 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158288 bases at least Q40

Consensus quality: 160068 bases at least Q30

Consensus quality: 160845 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 161363; sum-of-contigs

Quality coverage: 5.9 in Q20 bases; agarose-fp  
Quality coverage: 6.0 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3780: contig of 3780 bp in length  
3781 3880: gap of 100 bp  
3881 7396: contig of 3516 bp in length  
7397 7496: gap of 100 bp  
7497 14714: contig of 7218 bp in length  
14715 14814: gap of 100 bp  
14815 31890: contig of 17076 bp in length  
31891 31990: gap of 100 bp  
31991 54940: contig of 22950 bp in length  
54941 55040: gap of 100 bp  
55041 79944: contig of 24904 bp in length  
79945 80044: gap of 100 bp  
80045 104673: contig of 24629 bp in length  
104674 104773: gap of 100 bp  
104774 131627: contig of 26854 bp in length  
131628 131727: gap of 100 bp  
131728 162163: contig of 30436 bp in length.

FEATURES  
source

1. 162163  
/organism="Homo sapiens"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3"  
/clone="RP11-418M15"  
/clone\_1fb="RP11 Human Male BAC"  
1. 3780  
/note="assembly\_fragment"  
3781. 3880  
/estimated\_length=100  
3881. 7396  
/note="assembly\_fragment"  
clone end: 77  
vector side:right"  
7397. 7496  
/estimated\_length=100  
7497. 14714  
/note="assembly\_fragment"  
14715. 14814  
/estimated\_length=100  
14815. 31890  
/note="assembly\_fragment"  
31891. 31990  
/estimated\_length=100  
31991. 54940  
/note="assembly\_fragment"  
clone end: 5P6  
vector side:right"  
54941. 55040  
/estimated\_length=100  
55041. 79944  
/note="assembly\_fragment"  
79945. 80044  
/estimated\_length=100  
80045. 104673  
/note="assembly\_fragment"  
104674. 104773  
/estimated\_length=100  
104774. 131627  
/note="assembly\_fragment"  
131628. 131727  
/estimated\_length=100

```
misc_feature 131728..162163
/note="assembly_fragment"
ORIGIN
Query Match 25.3%; Score 505.2; DB 14; Length 162163;
Best Local Similarity 74.3%; Pred. No. 1.2e-69;
Matches 780; Conservative 0; Mismatches 238; Indels 32; Gaps 10;

Qy 965 AGGGCTAGATGATGACGATACCTTGAGATGATGCTGGGTTGGTTCATATCCACCACA 1024
Db 101233 AAGAGTTTACATATGCGGATACCTTGAGATGATGCTGGGTTGGTTCATATCCACCAGTGA 101292
Qy 1025 ATATATACAAATATGCAAGAGTGAATATGCAATATTAAGTATGACACAGCTTTTGGC 1084
Db 101293 CAAGAAATATATGCAATAA---ATATCAATATTAAGAGATGATACAAATATATATGT 101348
Qy 1085 TTCCAGTGCATATATTAAGTTTGTCTTAATATGATGATGCTGTTAAGTGTGA--A 1141
Db 101349 TTCCAGTGCATATATTAAGTTTGTCTTAATATGATGATGCTGTTAAGTGTGA--AAGCA 101408
Qy 1142 TAGGTGTATGCTAATAAATAACATACCTTAATTTTAAATGCTTATTAAT-AAAAAT 1200
Db 101409 CAGGTGTATGAAATACTATTTTACATACCTTAATGAAAAATGCTTTTGTGCTAAAAAAT 101468
Qy 1201 GCTAACATCATTTTGAGCATTCAGTGAATGATGATGATGATGATGATGATGATGATGATG 1260
Db 101469 GCTAACATCATTCAGCTTGAAGCTTGAAGATTCATATCTTTTGTGCTGAGGTTGATG 101528
Qy 1261 TTATGATGACTGA--TCGGGGGCTCAGGTGCTGAAGCTTGAAGGCTGCTGCGCACTTTC 1317
Db 101529 CTGCTGACGATCATCATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101588
Qy 1318 TTTAAACAAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1377
Db 101589 TTTAAACAAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 101648
Qy 1378 CTAGGTGTGATGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1436
Db 101649 GTAGGTGTGATGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 101708
Qy 1437 ATCAATCTCTGCAAAACC---TGCTGTGCTTTTAAACAACCTTAATATATATATTCGA 1493
Db 101709 ATCAATCTCTGCAAAACCCTGTTGCCACCTTAATATATATATATTCGA 101768
Qy 1494 ATCCATGTTGCTATTTCAACAATTTTCAAGTGTCTTCAACAAGATGATGATGATGATGATG 1553
Db 101769 ATCCATGTTGCTATTTCAACAATGTTCAACAATGATGATGATGATGATGATGATGATG 101828
Qy 1554 ATTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1613
Db 101829 AAGAAAGCACTTGG---TTTACTATCTGTAAGAGCACTTCTCATTCATGCAAGTTT 101884
Qy 1614 TATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1673
Db 101885 TATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 101939
Qy 1674 TTCTTGTGCTGTTTCAACAATCTGCTGCTTCTTCTTCAACAAGTGTGATGATGATGATG 1733
Db 101940 TTCTTGTGCTGTTTCAACAATCTGCTGCTTCTTCTTCAACAAGTGTGATGATGATGATG 101999
Qy 1734 CAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1793
Db 102000 AAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 102059
Qy 1794 GACCTC---CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1850
Db 102060 GACCTCCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 102119
Qy 1851 AAGGTTTTCATTTTACTAGTGCAGATGATGATGATGATGATGATGATGATGATGATGATG 1910
Db 102120 TGAGGTTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 102174
Qy 1911 GTTATAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1970
```

```
Db 102175 GCTATAGCTTACAAATGCTTATTAATATATAGGCTGAAGTCAAAATGACTCCT 102234
Qy 1971 TGATCATTTTCTGCAAAATATGATGATG 2000
Db 102235 TGATCATTTGCTGCAAGATGATGATG 102264

RESULT 39
AL929203/c 169424 bp DNA linear HTG 01-OCT-2002
LOCUS Homo sapiens chromosome 1 clone RP11-258u6.
DEFINITION AL929203
ACCESSION AL929203.3 GI:23476799
VERSION HTG: HTGS_PHASE2; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo
1 (bases 1 to 169424)
REFERENCE
1 Martin,S.
AUTHORS Direct Submission
TITLE Submitted (30-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgehire, CB10 1SA, UK. E-mail enquiries:
hunquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Oct 2, 2002 this sequence version replaced gi:23395907.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: hunquerry@sanger.ac.uk
----- Project Information
Center project name: ba258u6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 166627 bases at least Q40
Consensus quality: 166849 bases at least Q30
Consensus quality: 166945 bases at least Q20
Insert size: 169424; sum-of-coverage
Insert size: 172092; 4.2% error; agarose-fp
Quality coverage: 9.31x in Q20 bases; sum-of-coverage
Quality coverage: 9.16x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 169424: contig of 169424 bp in length.
FEATURES
source
1..169424
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-258u6"
/clone_id="RP11-11.1"
1..169424
misc_feature
/note="assembly_fragment:01428"
ORIGIN
Query Match 25.3%; Score 505.2; DB 14; Length 169424;
Best Local Similarity 74.3%; Pred. No. 1.2e-69;
Matches 780; Conservative 0; Mismatches 238; Indels 32; Gaps 10;

Qy 965 AGGGCTAGATGATGACGATACCTTGAGATGATGCTGGGTTGGTTCATATCCACCACA 1024
```



QY 1679 TTGCTGTTTACACACATCTGTGTTCTTCTCCATTTGAAGCTTGAACCTTCCAACT 1738  
 |||||  
 DB 105192 TTGCTATTTCTACACATATGACATGTTACTCTTCCACCTGAAGCTTTGAACCTTCAAACT 105251  
 |||||  
 QY 1739 CATCCAGAGGGTGGATGCACTTCTTCCAAATTCCTGTATAATTTATTTGACCT 1798  
 |||||  
 DB 105252 CATCCATGAGGGTGGATGCAATTTCTTCCAACTCTGTATATGTTGATATTTGACCT 105311  
 |||||  
 QY 1799 CC---CATGATCATGATGTTCTTATATGACACCTGGAATGTTGATCTTTCCAAAG 1855  
 |||||  
 DB 105312 CCTCATGAAATTAACAAAGATTCGTGAATGCACTTGAATGTTGATCTTTCCAAAG 105371  
 |||||  
 QY 1856 TTTTCATTTACTTGTGTCCAGATCCATCCAGAGATCCATCTTCAATGCCAGTTAT 1915  
 |||||  
 DB 105372 -TTTCATTTACTTGTGCCAGATCCAT---CAGAGATATCATTTCTATGACGACTAC 105425  
 |||||  
 QY 1916 AGCCTTATGAAATGATTTCTTCAATTAATAGCTTGAAGTTGAATTAATCTCTTGATC 1975  
 |||||  
 DB 105426 AGCCTTATGAAATGATTTCTTAAATGATTAACCTTGAAGTCAAAATTAATCTTGATC 105485  
 |||||  
 QY 1976 CATTTTCGCAAAATGATGTTGT 1999  
 |||||  
 DB 105486 CATGGCTGCAAGATGATGTTGT 105509  
 |||||

RESULT 41  
 LOCUS 175210 bp DNA linear PRI 25-JUN-2001  
 DEFINITION Human chromosome 14 DNA sequence BAC R-818K5 of library RPI-11  
 ACCESSION AL590328  
 VERSION AL590328.2 GI:14571668  
 KEYWORDS HTG; HTGS; ACTIVEFIN.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 175210)  
 Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,  
 Brocletier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,  
 Levy, M., Eckenberg, R., Bruls, T., deBardins, V., Cruaud, C.,  
 Gyapay, G., Saurin, W. and Weissenbach, J.  
 Sequencing of the human chromosome 14  
 Unpublished  
 2 (bases 1 to 175210)  
 Genoscope.  
 Direct Submission  
 Submitted (25-JUN-2001) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : segre@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr  
 On Jun 26, 2001 this sequence version replaced gi:13509295.  
 ----- Genome Center  
 Center: Genoscope / Centre National de Sequencage  
 Center code: GS  
 Web site: http://www.genoscope.cns.fr/  
 Contact: Segre@genoscope.cns.fr  
 -----  
 The following BAC sequence is oriented from the T7 to the SP6 end.  
 ----- Summary Statistics  
 Assembly program: Phrap, version 2.0  
 Quality coverage: 7.46x in Q20 bases; sum-of-contigs  
 -----  
 Overall quality chart :  
 Range : bases  
 0 : 6  
 1 - 9 : 52  
 10 - 19 : 268  
 20 - 29 : 576  
 30 - 39 : 1644  
 40 - 49 : 7554

50 - 59 : 9152  
 60 - 69 : 10935  
 70 - 79 : 21682  
 80 - 89 : 43918  
 90 - 99 : 79423  
 -----  
 Percentage of bases with a quality value >= 40 : 98 %  
 FEATURES  
 source location/Qualifiers  
 1..175210  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone="R-818K5"  
 /clone\_lib="RPI-11"  
 ORIGIN  
 Query Match 25.2%; Score 504; DB 8; Length 175210;  
 Best Local Similarity 75.0%; Pred. No. 1.8e-69;  
 Matches 784; Conservative 0; Mismatches 205; Indels 57; Gaps 10;  
 QY 973 ATAGTACAGCATACCTTGGAGATACGTGGTGTGTTCCATACCAACCAATATACA 1032  
 |||||  
 DB 164169 ACAGTATAGGATATCTTGGAGATCTGTAGTTCACTGCACACCTGCACTAAAA-- 164226  
 |||||  
 QY 1033 AATATGCAAGAGGATATACCAATTAAGAGTACACACAACTTTTGGCTCCAGT 1092  
 |||||  
 DB 164227 -----TGAAATTTGCAATTAAGAGGATGATGATGTTTGGTTCCAGT 164273  
 |||||  
 QY 1093 GCATTAATAAGTTTGTCTTATACCTACTGTAGTCTGTTAAGTGCATAGTGTATGT 1152  
 |||||  
 DB 164274 GCATTAATAAGTTTGTCTTAT-----ACTGTATGTCATTTAAGTGTGCATATGATGT 164328  
 |||||  
 QY 1153 CTAATAAAACACAT-----ACCTAATTTTAAATGCTTTTATTAATAAAATGCTAAC 1207  
 |||||  
 DB 164329 CTAATAAAAGTAAATTTACATCTTTAATTAATTAATTTAATAAAATGTTAATG 164388  
 |||||  
 QY 1208 ATCATTTAGCATTCAGAGATGTTATCTTTTGTCTGTGTGAAGTCTTTTCTTATGTA 1267  
 |||||  
 DB 164389 ATCATCTGAGTCTTCAAGCAATGTTAATTTTGTCTGTGTGAAGTCTTTTGTCTTATGT 164448  
 |||||  
 QY 1268 TGAATGAT-----CGGGGGTCAAGTGTGCTGAAGCTTGAAGGCTGTGGAGTTT 1316  
 |||||  
 DB 164449 TGAATGTTCTGAATGATCAAGGCTGTGTGTATGTAAGTGTGAAGTGTGTGGCAATTT 164508  
 |||||  
 QY 1317 CTTAAA-----ACAAAGTGAAGATTTGCATATATCACTTCTCTTCAATGAAGAT 1371  
 |||||  
 DB 164509 CTTAAAATTAAGCAACAGTGAAGTTGGCCACATGATGATCTTCTTCACAAAAAAG 164568  
 |||||  
 QY 1372 TTCTCTAGAGTGTGATGCTTTTGAATGCAATTTTATGACAGATGAATCTTTTGAAA 1431  
 |||||  
 DB 164569 TTCTCTAGAGTGTGATGCTTTTGAATGCAATTTTATGACAGATGAATCTTTTGAAA 164628  
 |||||  
 QY 1432 ATTGGA-TCAATCCCTCAAAACCTGCTCTTAAACAACCTAAGTAAATATATATTC 1490  
 |||||  
 DB 164629 ATTGATTCATTCCTCAAAACCTGCTCTTAAACAACCTAAGTAAATATATATTC 164687  
 |||||  
 QY 1491 TGAATCCATTTGTGATTTCAACAATTTTCAAGTGTCTTCAACAGAGATGATTCAT 1550  
 |||||  
 DB 164688 TGAATCCCTCAATTTCAACAAGTGTCTTCAACAGATCTTCAACAGAGATGATTCAT 164747  
 |||||  
 QY 1551 CTCATTTCTTGAATGATCTTTGCTGATCATTAAGAAATTTCTCATCTGTTCAAG 1610  
 |||||  
 DB 164748 CTC-----CAGAACTACTTTCTTGGCTCATGTGTGAGATCACTCTTATCTGTTCA 164803  
 |||||  
 QY 1611 TTTATCATGAGATTGACGAATACAGTCAATGCTTCAAGGCTCACTCACTTTAATTC 1670  
 |||||  
 DB 164804 TTGATCATGAGATGACGACATTCATTCATCTTCAAG-----ATCACTTCTAATTC 164858  
 |||||  
 QY 1671 CAGTTCTCTTGTCTTTTACACATCTGTGTTCTTCTTCATTTGAAGCTTTGAACCT 1730  
 |||||  
 DB 164859 TAGTTCTCTTGTCTTTTACACATCTGTGTTCTTCTTCATTTGAAGCTTTGAATTC 164918  
 |||||

QY 1731 CTCGAAGTACATGAGGGTGGAGATGCACTTCTCCAAATTCCTGTTAAATTTATAT 1790  
| | | | |  
Db 164919 CTCAAAGTCAGCCATGAGATAGATTAAGATTAACCTTCTCCAGATTCCTGTAATGTCCTTTT 164978  
| | | | |  
QY 1791 TTTGACCTCCATGATATGATATGTTTAAATGGCAGCTGGAATGGATCCCTTCA 1850  
| | | | |  
Db 164979 TTTGACCTCCATGATATGATATGTTTAAATGGCAGCTGGAATGGATCCCTTCA 165038  
| | | | |  
QY 1851 AAAGTTTCAATTTACTTAAGTCCAGATCCATCCAGAGATCCACTTTCAATGCCA 1910  
| | | | |  
Db 165039 AAAAGTTTCAATTTACTTTTCCAGATCCATCCAGAGATCCACTTTCAATGCCA 165093  
| | | | |  
QY 1911 GTTATAGCCTTATGAGATGATTTCTTCAATATATAGGCTGGAAGTTGAATTAATCTCT 1970  
| | | | |  
Db 165094 GTGATAGCCTTATGAGATGATTTTAAAGTATATACCTGGAAGTTGAATTAATCTCTCC 165153  
| | | | |  
QY 1971 TGATCCATTTTCTGCAAAATAGATCT 1996  
| | | | |  
Db 165154 TGATCCATGAGATGAGATGATGT 165179  
| | | | |

RESULT 42  
CNS01DSQ 200853 bp DNA linear PRI 26-SEP-2001  
LOCUS  
DEFINITION Human chromosome 14 DNA sequence BAC R-747H7 of library RPCI-11  
ACCESSION AL121839  
VERSION AL121839.3 GI:15796542  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo  
1 (bases 1 to 200853)  
Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,  
Brotier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,  
Levy, M., Beckenbery, R., Bruls, T., deBerardinis, V., Cruaud, C.,  
Gyday, G., Saurin, W. and Weissenbach, J.  
Sequencing of the human chromosome 14  
Unpublished  
2 (bases 1 to 200853)  
Genoscope.  
Direct Submission  
Submitted (26-SEP-2001) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
On Sep 27, 2001 this sequence version replaced gi:9955595.  
----- Genome Center  
Center: Genoscope / Centre National de Sequencage  
Center code: GS  
Web site: http://www.genoscope.cns.fr/  
Contact: Sequef@genoscope.cns.fr

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

The following BAC sequence is oriented from the T7 to the Sp6 end.  
Upstream BAC (overlapping the T7 end) : R-818K5 (AC=AL590328)  
Downstream BAC (overlapping the Sp6 end) : R-248B24 -----  
Summary Statistics  
Assembly Program: Phrap; version 2.0  
-----  
Overall quality chart :  
Range : bases  
0 : 1  
1 - 9 : 10  
10 - 19 : 37  
20 - 29 : 228  
30 - 39 : 3790  
40 - 49 : 12266  
50 - 59 : 14881  
60 - 69 : 23014  
70 - 79 : 49600  
80 - 89 : 49600

90 - 99 : 97026  
-----  
Percentage of bases with a quality value >= 40 : 99 %.  
FEATURES  
Location/Qualifiers  
1..200853  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="R-747H7"  
/clone\_id="RPCI-11"  
74706..74805  
/note="matching EMBL:R33632  
RHD:RH53920  
dbSTS:STS32121  
Identified using the e-PCR software (G. Schuler)"  
82445..82445  
/note="matching EMBL:R25802  
RHD:RH4322  
RHD:RH1302  
dbSTS:STS1025  
Identified using the e-PCR software (G. Schuler)"  
83496..83749  
/note="matching EMBL:H09564  
RHD:RH53633  
dbSTS:STS42877  
Identified using the e-PCR software (G. Schuler)"  
119346..119486  
/note="matching EMBL:G09073  
RHD:RH53763  
RHD:RH28183  
dbSTS:STS32298  
Identified using the e-PCR software (G. Schuler)"

ORIGIN  
Query Match 25.2%; Score 504; DB 8; Length 200853;  
Best Local Similarity 75.0%; Pred. No. 1,7e-69;  
Matches 784; Conservative 0; Mismatches 205; Indels 57; Gaps 10;  
QY 973 ATATGACGAGGATACCTTGGAGATCTGAGGTTTGGTTCATACCAACCAATATATCA 1032  
| | | | |  
Db 60151 ACAGTATGAGGATCTTGGAGATCTGAGGTTTGGTTCATACCAACCAATATATCA 60208  
| | | | |  
QY 1033 AATATGCAAGAGTGGATATCAATATAGATGACACAAAGCTTTGCTCCAGT 1092  
| | | | |  
Db 60209 -----TGAACATGCAATATAGAGGATGATGATTTTGGTTCCAGT 60255  
| | | | |  
QY 1093 GCATATTAAGTTTGGCTTATATCACTAGTCTGTTAAGTGGCAATAGTATATGT 1152  
| | | | |  
Db 60256 GCATATTAAGTTTGGCTTATATCACTAGTCTGTTAAGTGGCAATAGTATATGT 60310  
| | | | |  
QY 1153 CTAAAAAACAAT-----ACCTTAATTTAAATGCTTTATTAATAAATGCTAACA 1207  
| | | | |  
Db 60311 CTAAAAAAGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATG 60370  
| | | | |  
QY 1208 ATCATTTGAGCATTCAGTATGATTTTCTGCTGAGAGGCTTTTCTTATTTGA 1267  
| | | | |  
Db 60371 ATCATTTGAGCATTCAGTATGATTTTCTGCTGAGAGGCTTTTCTTATTTGA 60430  
| | | | |  
QY 1268 TGACTGAT-----CGGGGTCAAGTCTGAGCTTAAAGGTGGCTGGCAGTTT 1316  
| | | | |  
Db 60431 TGATGCTGCTGATATGATGAGGCTGGTGGTATGAAAGTTGAAGTGGCTGGCAGTTT 60490  
| | | | |  
QY 1317 CTTAAA-----ACAACAGTGAAGATTTGCAATATCAGTTCCTTCAAGAAAGT 1371  
| | | | |  
Db 60491 CTTAAAATTAAGCAACAGTGAAGTGGCCACATTAATGATCTTCTTCAAAAAAAG 60550  
| | | | |  
QY 1372 TTCTCTAGTGTGATGCTTTTGAATAGATTTTATGACAGTGAAGATCTTTTGA 1431  
| | | | |  
Db 60551 TTCTCTAGTGTGATGCTTTTGAATAGATTTTATGACAGTGAAGATCTTTTGA 60610  
| | | | |  
QY 1432 ATTGGA-TCAATCTCTCAACCCGCTCTGCTTTAACAACCTTAAGTTATATATTC 1490  
| | | | |

Db 60611 ATTGATTCATCTCTCAACCTGCTGTTCTTTC-AGCTAAGTTATGTAATATTC 60669  
 QY 1491 TGAATCCATTTGTTGATTTTCAACAATTTTCAAGTGTCTTCAACGAGTAGATTCAT 1550  
 Db 60670 TGAATCTCTGAATTTATTTTCAACAGTTCACAGATCTTCAACGAGTAGATTCAT 60729  
 QY 1551 CTCATTTCCGAGATGGAATCTTTGCTCATTAAGAAGAAATTCCTCATCTGTTCAAG 1610  
 Db 60730 CTC---CAGAAACTACTCTTCTTGTGCTCATGTGTGAATCAACTCTTATCTGTTCA 60785  
 QY 1611 TTTTATCATGAGATTCAGCAATATACGTCATGTTCTGAGGCTCCTCATCTTTAATTC 1670  
 Db 60786 TTGTATCATGAGATTCAGCAATATACGTCATGTTCTGAGGCTCCTCATCTTTAATTC 60840  
 QY 1671 CAGTCTCTTGTGCTTTTACCAATCTGTGTTCTTCTCTCATGAGTCTGAACT 1730  
 Db 60841 TTAGTCTCTGTTGCTTTTCCACACATCTGCAATTTACTCTCTGCGAAATTTGAAATC 60900  
 QY 1731 CTCCAAGTCATCAGGAGTTGGAATGCACTTCTTCAAAATTCCTGTTAATTTAAT 1790  
 Db 60901 CTCCAAGTCATCAGGAGTTGGAATGCACTTCTTCAAAATTCCTGTTAATTTAAT 60960  
 QY 1791 TTTGACCTCCATGATCATGATGATTTCTTAAATGCACTGGAATGTAATCTTTTCA 1850  
 Db 60961 TTTGACCTCCATGATCATGATGATTTCTTAAATGCACTGGAATGTAATCTTTTCA 61020  
 QY 1851 AAAGTTTTCATTTACTTATGTCAGATCCATCCAGAGATCCATTTCAATGCCA 1910  
 Db 61021 AAAGTTTTCATTTACTTATGTCAGATCCATCCAGAGATCCATTTCAATGCCA 61075  
 QY 1911 GTTATAGCCTTATGGAATGATTTTCTCAATATATAAGCTGGAATTTACTCTCT 1970  
 Db 61076 GTGATAGCCTTATGGAATGATTTTCTCAATATATAAGCTGGAATTTACTCTCT 61135  
 QY 1971 TGATCCATTTTCTGCAAAATAGATGT 1996  
 Db 61136 TGATCCATTTTCTGCAAAATAGATGT 61161  
 RESULT 43  
 AC004764 68048 bp DNA linear PRI 29-MAY-1998  
 LOCUS AC004764 Homo sapiens chromosome 5, p1 clone 255g5 (LBNL H61), complete  
 DEFINITION sequence.  
 AC004764 AC001063 AC002245 AC002246 L81711 L81712 L81884 L81885  
 AC004764.1 GI:3168619  
 VERSION L81886 L81887  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 68048)  
 AUTHORS Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M.,  
 Kader, K., Miguel, T., Miller, C., Pitluck, S., Pollard, M.,  
 Rojeski, H., Subramanian, S., and Martin, C. H.  
 TITLE Sequencing of human chromosome 5  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 68048)  
 AUTHORS Riche, D. O.  
 TITLE Large Scale Sequence Analysis and Annotation with the Sequence  
 Comparison Analysis (SCAN) System  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 68048)  
 AUTHORS Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M.,  
 Davis, C. A., Kader, K., Miguel, T., Pitluck, S., Pollard, M.,  
 Rojeski, H., Subramanian, S., and Martin, C. H.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAY-1998) Human Genome Center, DOE Joint Genome  
 Institute, Lawrence Berkeley National Laboratory, MS 74-157,  
 Berkeley, CA 94720, U.S.A.  
 COMMENT Sequence submitted by:

FEATURES DOE Joint Genome Institute.  
 Location/Qualifiers  
 1. 68048  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="5g"  
 /clone="255g5"  
 /note="LBNL H61"  
 repeat\_region complement(187. .491)  
 /rpt\_family="Alu"  
 517. .552  
 /note="(GT) 18"  
 /rpt\_type=tandem  
 /rpt\_unit="gt"  
 repeat\_region complement(548. .1010)  
 /rpt\_family="Alu"  
 repeat\_region complement(2267. .2592)  
 /rpt\_family="Alu"  
 repeat\_region complement(3203. .3434)  
 /note="GTAAT 2 excellent exon, frame 0"  
 /rpt\_family="Alu"  
 repeat\_region complement(5139. .5432)  
 /rpt\_family="Alu"  
 repeat\_region complement(5422. .5925)  
 /rpt\_family="Alu"  
 repeat\_region complement(5920. .6527)  
 /note="95% identity yx66g09.s1"  
 /db\_xref="dbEST:N23052"  
 6791. .7162  
 /rpt\_family="Alu"  
 repeat\_region complement(7294. .7362)  
 /rpt\_family="MIR42"  
 8033. .8105  
 /rpt\_family="MIR2"  
 8384. .8537  
 /note="GTAAT 2 excellent exon, frame 2"  
 complement(8593. .8928)  
 /rpt\_family="Alu"  
 repeat\_region complement(9004. .9331)  
 /rpt\_family="Alu"  
 9014. .9034  
 /note="(T) 21"  
 /rpt\_type=tandem  
 /rpt\_unit="t"  
 repeat\_region complement(9390. .9693)  
 /rpt\_family="Alu"  
 repeat\_region 9710. .9884  
 /rpt\_family="Alu"  
 repeat\_region complement(9901. .9985, 12737. .12991)  
 /standard\_name="ytc67a05"  
 /note="95% & 99% identity EST dbEST:R84804 & R85570"  
 10242. .10367  
 /note="GTAAT 2 excellent exon, frame 2"  
 10545. .10779  
 /rpt\_family="MIR46"  
 repeat\_region complement(10550. .10757)  
 /rpt\_family="MIR46"  
 11943. .12174  
 /rpt\_family="MIR42"  
 repeat\_region complement(12737. .12991)  
 /standard\_name="SHGC-30875"  
 /note="99% identity"  
 /db\_xref="dbSTS:G29311"  
 13888. .14069  
 /rpt\_family="Alu"  
 repeat\_region 14056. .14502  
 /rpt\_family="LI"  
 repeat\_region complement(14536. .14660)  
 /rpt\_family="MSTA"  
 14687. .14787  
 /rpt\_family="Alu"  
 repeat\_region complement(14871. .16160)



```
/rpt_family="MSTAR"
repeat_region complement(16201, .16573)
/rpt_family="MSTA"
repeat_region 16578, .16898
/rpt_family="L1MA5"
repeat_region 16970, .17136
/rpt_family="L1"
repeat_region complement(17157, .17470)
/rpt_family="Alu"
repeat_region complement(17596, .17863)
/rpt_family="Alu"
repeat_region 18183, .18338
/rpt_family="L1MB6"
repeat_region complement(18430, .18777)
/rpt_family="Alu"
repeat_region complement(18901, .19199)
/rpt_family="Alu"
repeat_region complement(19313, .19502)
/rpt_family="Alu"
repeat_region 19680, .19986
/rpt_family="Alu"
repeat_region 20240, .20302
/rpt_family="L1MB6"
repeat_region 20299, .20538
/rpt_family="Alu"
repeat_region 20955, .21167
/rpt_family="L1"
repeat_region 21807, .22046
/rpt_family="Alu"
repeat_region 22298, .23300
/rpt_family="L1"
repeat_region complement(23291, .23615)
/rpt_family="Alu"
repeat_region 23586, .23979
/rpt_family="L1MA2"
repeat_region 24086, .24388
/rpt_family="Alu"
repeat_region complement(24617, .25236)
/rpt_family="Alu"
repeat_region 24925, .24949
/note="(T) 21"
/rpt_type=tandem
/rpt_unit="c"
25335, .26483
/rpt_family="L1"
repeat_region complement(26476, .28579)
/rpt_family="T19ger1"
repeat_region complement(28589, .29033)
/rpt_family="MER4"
repeat_region complement(29028, .29221)
/rpt_family="T19ger1"
repeat_region 29445, .30152
/rpt_family="L1"
repeat_region 30319, .30630
/rpt_family="Alu"
repeat_region complement(30604, .30887)
/rpt_family="RBL10C4"
repeat_region 30896, .31218
/rpt_family="Alu"
repeat_region 31213, .31918
/rpt_family="L1"
repeat_region 32291, .32625
/rpt_family="Alu"
repeat_region complement(33507, .33877)
/rpt_family="L1"
repeat_region complement(35113, .35272)
/rpt_family="MIR"
repeat_region complement(36395, .42569)
/rpt_family="L1"
repeat_region complement(40993, .41144)
/rpt_family="MER25"
45011, .45337
/rpt_family="Alu"
```

```
repeat_region complement(45637, .45703)
/rpt_family="MIR"
misc_feature 46384, .46629
/note="GRAIL 2 excellent exon, frame 0"
misc_feature 47310, .47464
/note="GRAIL 2 excellent exon, frame 2"
misc_feature 47840, .47963
/note="GRAIL 2 excellent exon, frame 1"
misc_feature 48883, .49040
/note="GRAIL 2 excellent exon, frame 0"
prim_transcript complement(join(48972, .49041,50028, .50217))
/standard_name="hg41hl1"
/note="99% identity dbEST:AA527520"
misc_feature 50031, .50133
/note="GRAIL 2 excellent exon, frame 0"
repeat_region complement(52346, .52761)
/rpt_family="THE1"
repeat_region 54686, .54836
/rpt_family="Alu"
repeat_region 56010, .56375
/rpt_family="Alu"
prim_transcript join(57722, .57914,60189, .60588)
```

Query Match 25.2%; Score 503.4; DB 8; Length 68048;  
Best Local Similarity 77.7%; Pred. No. 3.2e-69;  
Matches 736; Conservative 0; Mismatches 171; Indels 40; Gaps 9;

```
QY 1051 ATCAATTAAGTGTGACACAAAGTCTTTGGCTTCCAGTGCATATAAAGTTTGCT 1110
DB 26515 ACCCAATTAAGTGTGACACAAAGTCTTTGGCTTCCAGTGCATATAAAGTTTGCT 26574
QY 1111 TATACACAGTGTGACACAAAGTCTTTGGCTTCCAGTGCATATAAAGTTTGCT 1169
DB 26575 TACATTATACAGTGTGACACAAAGTCTTTGGCTTCCAGTGCATATAAAGTTTGCT 26634
QY 1170 TTAATTTTAAATGCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1229
DB 26635 TTAATTTTAAATGCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 26694
QY 1220 TGTATCTTTTGGCTGTGAGGAGTCTTTTGTATTAATTAATTAATTAATTAATTA 1278
DB 26695 CATATCTTTTGGCTGTGAGGAGTCTTTTGTATTAATTAATTAATTAATTAATTA 26754
QY 1279 GGTGACAGTGTGAGGAGTCTTTTGTATTAATTAATTAATTAATTAATTAATTAAT 1333
DB 26755 ATGATGAGTGTGAGGAGTCTTTTGTATTAATTAATTAATTAATTAATTAATTAAT 26814
QY 1334 GATTGCAATATCAATGTTGACTCTTCTTCATGAAGATTTCTCTAGTGTGATGCTT 1393
DB 26815 GTTTCACATCAATGTTGACTCTTCTTCATGAAGATTTCTCTAGTGTGATGCTT 26874
QY 1394 TTTGATAGATTTTATGACAGTATGACTCTTGTAAATTTGA-TCAATCTCTCAAC 1452
DB 26875 TTTGTTGAATTTTATGACAGTATGACTCTTGTAAATTTGA-TCAATCTCTCAAC 26934
QY 1453 CCTG-CCTGCTTTAACAACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1511
DB 26935 CCTGCACCTCTTTATGACAGTATGACTCTTGTAAATTTGA-TCAATCTCTCAAC 26994
QY 1512 AACAACTTTTCAAGTGTCTTCAACGAGTATGACTCTATTTCTGAGATGGAATC 1571
DB 26995 AACAACTTTTCAAGTGTCTTCCGACGAGTATGACTCTCAAGAACCAAGTT---TC 27050
QY 1572 TTTGCTCATCCATGAAGAATTTCTCATCTGTTCAAGTTTATCATGAGATGAGCA 1631
DB 27051 TTTGCTCATCCATGAAGAATTTCTCATCTGTTCAAGTTTATCATGAGATGAGCA 27110
QY 1632 ATACAGTATGCTTCAAGGCTCACTTCACTTTATTTCAAGTCTCTGCTGTTTCA 1691
DB 27111 GTTTAGTCACTTTCAAG-----CTCATTTCTAATTTAATTTCACTTGTATTTCCAG 27165
QY 1692 CACATCTGTGTTCTTCTTCATGAAGTCTTGAACCTTCAAGTATTCATGAGGT 1751
```

Db 27166 AACATCTGCAATTACTTCTCCAGTAGAGTCTTGAAACCCCTCAAGTCATCCAGAGGT 27225  
Qy 1752 TGGATCGACTTCTTCCAAATTCCTGTATATATTAATTTGA---CTCCCAATGAATC 1808  
Db 27226 TGGAAATCAACTCTTCCAAATTCCTGTATATTAATTTGAATCTCTTCCTGTAGTC 27285  
Qy 1809 ATGAATGTTCTTAATGAGCAGCTGGAATGAGTAATCTCTTCCAAAAGGTTTCAATTACT 1868  
Db 27286 CTGAATGTTCTTAATGAGCAGCTGGAATGAGTAATCTCTTCCAAAAGGTTTCAATTACT 27345  
Qy 1869 TAGTCAGATTCATCCATCCAGAGATCCATCTTCAATGCCATTAATGCTTATGGAAT 1928  
Db 27346 TTACCCAGAGCCCATCC-----GAAAGATCATATGTCAGCTATAGCCTTATGAAT 27396  
Qy 1929 GTATTTCTTCAATTAATGAGCTTGAAGTTGAATTAATCTCTTGATC 1975  
Db 27397 GTATTTCTTCAATTAATGAGCTTGAAGTTGAATTAATCTCTTGATC 27443

RESULT 44  
AC008670/c 126034 bp DNA linear PRI 28-JAN-2003  
LOCUS Homo sapiens chromosome 5 clone CTB-3601, complete sequence.  
DEFINITION AC008670  
AC008670.6 GI:27923608  
VERSION  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eumetazoa; Eutheria; Euarthropods; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE 1 (bases 1 to 126034)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 126034)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Direct Submission  
REFERENCE 3 (bases 1 to 126034)  
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
REFERENCE 4 (bases 1 to 126034)  
AUTHORS Submitted (22-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
REFERENCE 5 (bases 1 to 126034)  
AUTHORS Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
REFERENCE 6 (bases 1 to 126034)  
AUTHORS Submitted (28-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
JOURNAL On Jan 28, 2003 this sequence version replaced gi:15187197.  
COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 2.1.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 133kb). It is clipped at the overlap with AC006077. The  
number of bases overlapped is 36042.  
Location/Qualifiers  
1. 126034  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTB-3601"

ORIGIN

Query Match 25.2%; Score 503.4; DB 8; Length 126034;  
Best Local Similarity 77.7%; Pred No. 2.5e-69;  
Matches 736; Conservative 0; Mismatches 171; Indels 40; Gaps 9;

Qy 1051 ATCAATTAAGTGAAGTCAACAAGCTTTGGCTTCCAGTGCATATTAAGTTTGGT 1110  
Db 22210 ACCACATTAAGTGAAGTCAACAAGTTTGGTTCCTGATATTAAGTTTGGT 22151  
Qy 1111 TATATCACTAGTGTGTGTAAGTGTGAATGTGTATGTC-TAAAAAACACATCC 1169  
Db 22150 TACATTAATCTAGTGTGTGTAAGTGTGAATGTGTATGTC-TAAAAAACATTAAT 22091  
Qy 1170 TTAATTTAAATGCTTTTAACTTAAATGTAATGTAATTAATGATTAAGTGAAGT 1229  
Db 22090 TTAATTTAAATGCTTTTAACTTAAATGTAATGTAATTAATGATTAAGTGAAGT 22031  
Qy 1230 TGAATCTTTTGGTGGAGAGTCTTTTCTTATTAATGATGATGAT-----CGG 1278  
Db 22030 CATTAATCTTTTGGTGGAGAGTCTTTTCTTATTAATGATGATGATGATGATCAAG 21971  
Qy 1279 GAGTCAGTGTGAGAGCTTAAGGCTGTGTGCAAGTTCTTAA-----ACAAGTGA 1333  
Db 21970 ATGATGTTGTGAGAGCTTAAGGCTGTGTGCAAGTTCTTAAATTAAGACAAAGAA 21911  
Qy 1334 GATGCAATTCAGTGAATCTTCTTCTTATTAATTAATTAATTAATTAATTAATTAAT 1393  
Db 21910 GTTGTGCAATTCAGTGAATCTTCTTCTTATTAATTAATTAATTAATTAATTAATTAAT 21851  
Qy 1394 TTGATGATGATTTATGACAGTAGAATCTTTTGAATTAATGA-TCAATCTCTCAAC 1452  
Db 21850 TTGATGATGATTTATGACAGTAGAATCTTTTGAATTAATGA-TCAATCTCTCAAC 21791  
Qy 1453 CCTG-CTGTGCTTAAACAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 1511  
Db 21790 CCTGCACTGCTTATTAACAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 21731  
Qy 1512 AACATTTTCAAGTGTCTTCAACAGAGTAGATTCATCTCAATTTCTTCAAGTGAATC 1571  
Db 21730 AACATTTTCAAGTGTCTTCAACAGAGTAGATTCATCTCAATTTCTTCAAGTGAATC 21675  
Qy 1572 TTTGCTCATCCATTAAGAAATTCCTCATCTGTTCAAGTTTATCATGAGATTCAGCA 1631  
Db 21674 TTTGCTCATCCATTAAGAAATTCCTCATCTGTTCAAGTTTATCATGAGATTCAGCA 21615  
Qy 1632 ATACAGTCAATGTTCTTCAAGGCTCACTTCACTTTAATTCAGTCTCTGTGTTCTAC 1691  
Db 21614 GTTATGTCATCTTCAAG-----CTCCATTTCTAATTCATGTTCACTTGTATTTCCAC 21560  
Qy 1692 CACATCTGTGTTCTTCTTCTTCAAGTGAATTCGAACCTTCAAGTCAATTCAGAGGT 1751  
Db 21559 AACATCTGCAATTAATCTTCTTCAAGTGAATTCGAACCTTCAAGTCAATTCAGAGGT 21500  
Qy 1752 TGGAAATCGACTTCTTCCAAATTCCTGTATTAATTAATTTGA---CTCCCAATGAATC 1808  
Db 21499 TGGAAATCGACTTCTTCCAAATTCCTGTATTAATTAATTTGAATCTTCTCGTAGTC 21440  
Qy 1809 ATGAATGTTCTTAATGAGCAGCTGGAATGAGTAATCTTCCAAAAGGTTTCAATTACT 1868  
Db 21439 CTGAATGTTCTTAATGAGCAGCTGGAATGAGTAATCTTCCAAAAGGTTTCAATTACT 21380  
Qy 1869 TAGTCAGATTCATCCATCCAGAGATCCATCTTCAATGCCATTAATGCTTATGGAAT 1928  
Db 21379 TTACCCAGAGCCCATCC-----GAAAGATCATATGTCAGCTATAGCCTTATGAAT 21329  
Qy 1929 GTATTTCTTCAATTAATGAGCTTGAAGTTGAATTAATCTCTTGATC 1975  
Db 21328 GTATTTCTTCAATTAATGAGCTTGAAGTTGAATTAATCTCTTGATC 21282

RESULT 45  
AL357520 178942 bp DNA linear PRI 18-MAY-2005  
LOCUS

```

DEFINITION      Human DNA sequence from clone RP11-288D15 on chromosome 10 Containing
                  a pseudogene similar to part of glycogen synthase kinase 3 and a
                  ubiquitin-conjugating enzyme pseudogene, complete sequence.
ACCESSION       AL357520
VERSION         AL357520.12 GI:15620611
KEYWORDS        HTG; ubiqutin-conjugating.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
                Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Homnidae; Homo.
REFERENCE       1 (bases 1 to 178942)
AUTHORS         Pearce A.
JOURNAL         Direct Submission
TITLE           Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT         Clone request(s): clonerequest@sanger.ac.uk
                On Sep 14, 2001 this sequence version replaced gi:14485342.
                The following abbreviations are used to associate primary accession
                numbers given in the feature table with their source databases:
                Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WormPEP; Information
                on the WormPEP database can be found at
                http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
                was generated from part of bacterial clone configs of human
                chromosome 10, constructed by the Sanger Centre Chromosome 10
                Mapping Group. Further information can be found at
                http://www.sanger.ac.uk/HGP/Chr10
                RP11-288D15 is from the library RPECI-11.1 constructed by the group
                of Pieter de Jong. For further details see
                http://www.chori.org/bacpac/home.htm
                VECTOR: pBACe3.6
                -----Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: vegas@sanger.ac.uk
                -----
                This sequence was finished as follows unless otherwise noted: all
                regions were either double-stranded or sequenced with an alternate
                chemistry or covered by high quality data (i.e., phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats; all regions were covered by at least
                one subclone, and the assembly was confirmed by restriction digest,
                except on the rare occasion of the clone being a YAC.
FEATURES             location/Qualifiers
     source          1..178942
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="caxon:9606"
                     /chromosome="10"
                     /clone="RP11-288D15"
                     /clone_1fb="RPCI-11.1"
     misc_feature    1
                     /note="Clone left_end: RP11-288D15"
     misc_feature    43725..43746
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    54175
                     /note="Clone right_end: RP11-576H16"
     gene            66599..67051
                     /locus_tag="RP11-288D15.1-001"
                     /pseudo
     CDS              66599..67051
                     /locus_tag="RP11-288D15.1-001"
                     /locus_tag="RP11-288D15.1-001"
                     /locus_tag="RP11-288D15.1-001"
                     /note="Match: proteins: AAFA5801 AAH27984 AAM52705
                     AAN09082 AAN09083 P18265 P18431 P49840 Q6MRP7 Q91757
                     Q91BQ3 Q9NL43 Q9YH61"
                     /pseudo
                     /codon_start=1
                     /product="pseudogene similar to part of glycogen synthase
                     kinase 3"
     misc_feature    90945
                     /note="Clone left_end: RP11-498A5"

```

misc_feature	104374	/note="C1one_right_end: RP11-366F13"
gene	join(177553_177735,178054_178268)	
	/locus_tag="RP11-288D15.2-001"	
CDS	/pseudo	
	join(177553_177735,178054_178268)	
	/locus_tag="RP11-288D15.2-001"	
	/note="match: proteins: Q13403 Q13404 Q15819 Q8K2V7 Q8VB85 Q98079 Q96H34 Q9CVD7 Q9CZY3 Q9ZDM8 Q9ER18 Q9EZM1 Q9H454 Q9N1C3 Q9SVD7 Q9UM49 Q9VKL1"	
	/pseudo	
	/codon_start=1	
	/product="ubiquitin-conjugating enzyme pseudogene"	
misc_feature	178942	
	/note="C1one_right_end: RP11-288D15"	
ORIGIN		
Query Match	25.2%; Score 503.4; DB 8; Length 178942;	
Best Local Similarity	73.6%; Pred. No. 2,2e-69;	
Matches 766; Conservative	0; Mismatches 231; Indels 44; Gaps 8;	
QY	974 TAGACAGGCAATCCTTGAGGATACGTCGTGGTTGGTCCATACCAACCAATATACAA	1033
Db	49673 TATTATAGGTAACCAATAGAGATATGTGATCACTTCCAAACCAATATATAG----	49728
QY	1034 ATATGCAAGAGGAGATATACAAATATAGAGTCAACAAGTCTTTGGCTCCACAGT	1093
Db	49729 -----GTGGATATCATATATTAAGAGGTTAAATATTTTGGTTCCAGTA	49777
QY	1094 CAAATAAAAGTTTGGCTTATATACACCTGATCTGTTAAGTGCATATAGTTATATGC	1153
Db	49778 CCAACAAGATTATATTTACATCATATCATAGCTTATTAAGTGCAACAATATATATATGC	49837
QY	1154 TAAATAA-ACACATCTTATTTTAAATNGCTTATTACTTAAATAAATCTAACATAT	1212
Db	49838 TAAATAAATATTAATATCTTATATTAATAAATACTTATATCTTAAACCTGCTATATATCT	49897
QY	1213 TTGAGCAATCAGTAGAGTTGTAATCTTTTGTCTGTGGAGAGTCTTTTCTTATGTAGACT	1272
Db	49898 CGAGTCTTCAAGTAAATTTGTAATCTTTTGTCCACGAGAGGTCCTTGATGTTGATNG	49957
QY	1273 GAT-----CGGGGGTCAAGGTGCTGAAGCTTAAGGTTGGCTGTGCACTTTCTTAA	1321
Db	49958 GCTGCTGACTGATCAGGGGTGGTGGTTGCTGAAGATTTGAGGTGCTCGAATAATTTCTTAA	50017
QY	1322 AACAAAGTAGAAGATTGCATATACAGTACCTTCCATCANGAAGATTCTCTCATAG	1381
Db	50018 GACAAACATGAGTTCCACCAATGATGATTTCTTACTTATGAAAGATTATCTAATAG	50077
QY	1382 TGTGTGATNGCTTTTGTATGACATTTTATGACAGTAAACTCTTTGAAATAATGGA-TCA	1440
Db	50078 CGTGTATGTTGTCTCATACATTTTACCACAGAGAAAGTCCTTTAAATTTGGAGTCA	50137
QY	1441 ATCTCTCAAACTCTGCTGTGCTTTAACAACCTAAGTAAATATATATTTCTGAATCAT	1500
Db	50138 ATCTTCTCATACCAATGCAATGCTTATTAACAATTAAGTAAATATCTTAAACCTTT	50197
QY	1501 GTTGTCAATTTCAACAATTTCAACAGTCTTCAACGAGTAGATTCATCATCATTTCTCT	1560
Db	50198 GTTGTCAATTTCAATAAATGTTCAACAGATCTTCACTGAGGTAGATTTCAATCTCA	50253
QY	1561 GAGATGGAATCTTGTGCTATCCATTAAGAAGAAATTTCTCATCTGTCAAGTTTATCAG	1620
Db	50254 AAACACTTCTTTGCTTATTCGTAAGGGCACTCATCTGTTGGAAGTTTGTCAAG	50313
QY	1621 AGATTGACAAATACATCATGCTTCAAGGCTCACTCACTTTAATTCAGATTCTCTT	1680
Db	50314 AGATTGACAAATTCAGTCAAGATTTTCAAGTCTTATTTCCAAATTTAGTTTC-----TCT	50367
QY	1681 GCTGTTTCTAACCAATCTGTGGTCTCTTCTCCATGAAGTCAAGTCAACCTTCCAAGTCA	1740
Db	50368 CGATTTTGGACCAATCTGAGTGTGTTTCTTCCACGAACTTTGAACTTCTCAAAAGTCA	50427

```

Qy 1741 TCCATGAGGAGTGGAAATCGACTTCTTCCAAATTCCTTAATAATTATATTGACCTCC 1800
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 50428 TCCATGAGATTTAGATAAATCTTTCCAAATCTTGTAAATGATATTTTACCTTC 50487
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1801 CATGAATCANTGAATGTTCTTAATGCACTGGAAATGGTGAATCCTTCCAAAAGTTTTC 1860
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 50488 CATGAATCANTGAATGTTCTTAATGCACTGGAAATGGTGAATCCTTCCAAAAGTTTTC 50547
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1861 AATTTACTAGTCCATGCAATCCATCCAGAGGATCCACTTGCATGCCAGTTATAGCC- 1919
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 50548 AACTTTCTTTTCCAGATTCAT-----CAGAGGAATCAATTTTATGGCAGCTATAGCCT 50602
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1920 TTATGGAATGTATTTCTTCATATATAGGCTTGAAGTTGAATTAATCTCTTGATCCATT 1979
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 50603 TTATGGAATGTATTTCTTCATATATAGGCTTGAAGTTGAATTAATCTCTTGATCCATT 50662
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1980 TTCTGCAAAAATAGATGTGTG 2000
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 50663 AGCTGCAGAGATGATGTGTG 50683
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: January 22, 2006, 13:37:04  
 UOB time : 6837.28 secs

This Page Blank (upside)



c 93 403.4 20.2 15809 4 AAK68321 Human Imm  
94 401 20.1 106544 13 ABD32669 Human can  
95 399 20.0 76600 12 AD079402 regl  
96 396.2 19.8 101569 12 ADP13274 Renal cel  
97 396.2 19.8 147309 6 ABK49450 Human tra  
98 394.4 19.7 117813 14 ADX07540 Cyclin-de  
99 393.8 19.7 110000 10 ADE11169\_1 Continuation (2 of  
100 393 19.7 112190 4 AAH44801 Human GPC

## ALIGNMENTS

## RESULT 1

ID AAV83939 standard; DNA; 80595 BP.

AC AAV83939;

DT 03-MAR-1999 (first entry)

DE HC-contig derived from normal human chromosome 10q25.2 region.

XX Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;

XX neocentromere; replication; extra-chromosomal element; segregation;

XX cell division; artificial chromosome; gene therapy; mardel(10);

XX human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.

OS Homo sapiens.

XX WO9851790-A1.

XX 19-NOV-1998.

XX 13-MAY-1998; 98WO-AU000352.

XX 13-MAY-1997; 97AU-00006784.

XX 26-AUG-1997; 97AU-00008791.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Choo K, Du Sart D, Cancilla MR;

XX MPI; 1999-009773/01.

XX New isolated nucleic acid comprising neocentromere sequences from

XX eukaryotic chromosome - used to produce replicable, segregating

XX artificial chromosomes that can carry large amounts of DNA for gene

XX therapy.

XX Claim 8; Fig 6; 540bp; English.

XX The present sequence represents the HC-contig derived from normal human

XX chromosome 10, 10q25.2 region. This region can be naturally mutated to

XX produce an unusual chromosomal marker referred to as mardel(10). The

XX mardel(10) marker is mitotically stable and contains a functional

XX neocentromere at a location regarded as non-centromeric. This

XX neocentromere maps to q25.2 on chromosome 10. The specification describes

XX nucleic acid sequences derived from a eukaryotic chromosome, including a

XX neocentromere or its functional derivative or hybrid, that are able, in a

XX compatible cell, of replicating, acting as extra-chromosomal element and

XX segregating during cell division. The sequences can be used to construct

XX artificial chromosomes for use in gene therapy comprising a replicable,

XX segregating nucleic acid that confers a specific phenotype on cells.

XX Human artificial chromosomes can propagate in human cells and carry large

XX amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,

XX they are not mutagenic. The artificial chromosomes are also useful for

XX generation of transgenic plants and animals, in production of proteins

XX and to make diagnostic reagents, e.g. for expression of cytokines,

XX receptors and growth factors, or to increase the copy number of a gene in

XX a cell. The constructs may also be used for functional and structural

XX analysis of chromosomes

SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

Query Match 100.0%; Score 2000; DB 2; Length 80595;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCCTGCTCCAGGCTCCCAAGTAGGTTACAGGTGCACGACCAACGTCGA 60

Db 1 GAATTCCTGCTCCAGGCTCCCAAGTAGGTTACAGGTGCACGACCAACGTCGA 60

QY 61 GCTAATTTTGTATTTTGTAGAGACGGGGTTTACCGGTGTTGACGCTGATCAAA 120

Db 61 GCTAATTTTGTATTTTGTAGAGACGGGGTTTACCGGTGTTGACGCTGATCAAA 120

QY 121 CTCCTGACCTCAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180

Db 121 CTCCTGACCTCAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180

QY 181 AGTCACCGACCCAGCCCTTCTTTCAGTTCTATCACCCTTTTGTCTATTTGTAG 240

Db 181 AGTCACCGACCCAGCCCTTCTTTCAGTTCTATCACCCTTTTGTCTATTTGTAG 240

QY 241 AGCTTATTTATTTAGGGGCAATACATTTAAATGTATGCTATGATGATGATCT 300

Db 241 AGCTTATTTATTTAGGGGCAATACATTTAAATGTATGCTATGATGATGATCT 300

QY 301 GTCATTATGAATGCTGATTCATTCCTGATAGATATTCCTTTCTAAATATTTTCTG 360

Db 301 GTCATTATGAATGCTGATTCATTCCTGATAGATATTCCTTTCTAAATATTTTCTG 360

QY 361 AATGTCCTGCTATTAACATACACCTGCTGCTTTTAAATATGATTTTATAGCTATA 420

Db 361 AATGTCCTGCTATTAACATACACCTGCTGCTTTTAAATATGATTTTATAGCTATA 420

QY 421 TATTTTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 480

Db 421 TATTTTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 480

QY 481 GTGCTTATAGGACATATATCTGGGCTTGATGATATTAATTAATGATATCTCAAC 540

Db 481 GTGCTTATAGGACATATATCTGGGCTTGATGATATTAATTAATGATATCTCAAC 540

QY 541 CTTTTTGTGAGTGTATAGGCAATTAATTAATTAATTAATTAATTAATTAATTA 600

Db 541 CTTTTTGTGAGTGTATAGGCAATTAATTAATTAATTAATTAATTAATTAATTA 600

QY 601 GCTATACCATCTTTTCAATTTGTTTATATGATGACCATCTTTTCAATTTTCAAT 660

Db 601 GCTATACCATCTTTTCAATTTGTTTATATGATGACCATCTTTTCAATTTTCAAT 660

QY 661 CTTTGACCATTTTCTTATAGTACGAACTTTTGTATTTGATTTCAATTAATTCAT 720

Db 661 CTTTGACCATTTTCTTATAGTACGAACTTTTGTATTTGATTTCAATTAATTCAT 720

QY 721 TTTAGTATACCTTTAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 780

Db 721 TTTAGTATACCTTTAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 780

QY 781 TTTACTTATACAGATTTACTTCAATTAATTTTACAGGCAATGATAGTATAGAAC 840

Db 781 TTTACTTATACAGATTTACTTCAATTAATTTTACAGGCAATGATAGTATAGAAC 840

QY 841 CTTACAGAGTATTTTCAATTTTCTTCTATTTTATGCTATGATGATGATGATGAT 900

Db 841 CTTACAGAGTATTTTCAATTTTCTTCTATTTTATGCTATGATGATGATGATGAT 900

QY 901 AGGTTTGTGTTTGTGTTTGTATTAATTTTCTTTTCTTTTCTTTTCTTTTCTTT 960

Db 901 AGGTTTGTGTTTGTGTTTGTATTAATTTTCTTTTCTTTTCTTTTCTTTTCTTT 960

QY 961 GTTAAAGGCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020

Db 961 GTTAAAGGCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020



QY 1021 CACAAATTAATACAAATATGCAAGAGTATATCATATTAAGAGTCAACAAGCTTT 1080  
DB 1021 CACAAATTAATACAAATATGCAAGAGTATATCATATTAAGAGTCAACAAGCTTT 1080  
QY 1081 TGGCTTCCAGTGCATATATAAAGTTTGGTTATCTACACTGAGTCTGTAAAGTGCA 1140  
DB 1081 TGGCTTCCAGTGCATATATAAAGTTTGGTTATCTACACTGAGTCTGTAAAGTGCA 1140  
QY 1141 AATAGTTTATGCTATAAATAAACAATACCTTAATTTAAATGCTTATTAATTAATAAT 1200  
DB 1141 AATAGTTTATGCTATAAATAAACAATACCTTAATTTAAATGCTTATTAATTAATAAT 1200  
QY 1201 GCTAACATCATTTGAGCATTCAGTGTATATCTTTTGTGTGAGAGTCTTTTC 1260  
DB 1201 GCTAACATCATTTGAGCATTCAGTGTATATCTTTTGTGTGAGAGTCTTTTC 1260  
QY 1261 TTATGATGATGATCGGGGGTCAAGTGTGAAGCTTAGGGTGGCTGTGGAGTTCTTA 1320  
DB 1261 TTATGATGATGATCGGGGGTCAAGTGTGAAGCTTAGGGTGGCTGTGGAGTTCTTA 1320  
QY 1321 AAACAACAGTGAAGATTCGAATATCAGTTGACTCTTCTTATGAAGAATTTCTCTTA 1380  
DB 1321 AAACAACAGTGAAGATTCGAATATCAGTTGACTCTTCTTATGAAGAATTTCTCTTA 1380  
QY 1381 GGTGTGATGCTTTTATGAGATTTTATGACAGTGAAGACTTTGAAATTTGATCA 1440  
DB 1381 GGTGTGATGCTTTTATGAGATTTTATGACAGTGAAGACTTTGAAATTTGATCA 1440  
QY 1441 ATCTCTCAAACTGCTCTGCTTTTAACTAAGTTAATTAATTTCTGAATTCAT 1500  
DB 1441 ATCTCTCAAACTGCTCTGCTTTTAACTAAGTTAATTAATTTCTGAATTCAT 1500  
QY 1501 GTTGTCAATTTCAACAATTTTCAAGTGTCTTCAAGAGTGAATTCATCATTTCT 1560  
DB 1501 GTTGTCAATTTCAACAATTTTCAAGTGTCTTCAAGAGTGAATTCATCATTTCT 1560  
QY 1561 GAGATGSAATCTTTGCTCATCAATAAGAAATTCCTGATCTGTTCAAGTTTATCAG 1620  
DB 1561 GAGATGSAATCTTTGCTCATCAATAAGAAATTCCTGATCTGTTCAAGTTTATCAG 1620  
QY 1621 AGATTGCAAGATACAGTATCTTCAAGGCTCACTTCACTTTAATTCAGAGTTCTCT 1680  
DB 1621 AGATTGCAAGATACAGTATCTTCAAGGCTCACTTCACTTTAATTCAGAGTTCTCT 1680  
QY 1681 GCTGTTTCTACCAATCTGTGTCTCTTCCATTTGAAGTCTTGAACCTCTCAAGTCA 1740  
DB 1681 GCTGTTTCTACCAATCTGTGTCTCTTCCATTTGAAGTCTTGAACCTCTCAAGTCA 1740  
QY 1741 TCCATGAGGTTGSAATGCACTTCCAAATTCCTGTTAATTAATTTTGAAGCTCC 1800  
DB 1741 TCCATGAGGTTGSAATGCACTTCCAAATTCCTGTTAATTAATTTTGAAGCTCC 1800  
QY 1801 CATGATCATGATATGTTCTTAATGCACTGSAATGGAATCTTTCCAAAGGTTTTC 1860  
DB 1801 CATGATCATGATATGTTCTTAATGCACTGSAATGGAATCTTTCCAAAGGTTTTC 1860  
QY 1861 AATTACTTAGTTCAGATTCATTCAGAGATTCACCTTTCAATGCCAGTTATAGCT 1920  
DB 1861 AATTACTTAGTTCAGATTCATTCAGAGATTCACCTTTCAATGCCAGTTATAGCT 1920  
QY 1921 TATGGAATGATTTCTTCAATAATAAGGTTGAAGTTGAATTAATCTCTTATATCATTT 1980  
DB 1921 TATGGAATGATTTCTTCAATAATAAGGTTGAAGTTGAATTAATCTCTTATATCATTT 1980  
QY 1981 TCTGCAAAATAGATTTGTG 2000  
DB 1981 TCTGCAAAATAGATTTGTG 2000

RESULT 2  
AAV83940  
ID AAV83940 standard; DNA; 80240 BP.

XX AAV83940;  
AC 03-MAR-1999 (first entry)  
XX NC-config derived from mardel(10) on chromosome 10q25.2.  
DE Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;  
XX neocentromere; replication; extra-chromosomal element; segregation;  
KW cell division; artificial chromosome; gene therapy; mardel(10);  
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.  
XX Homo sapiens.  
OS WO9851790-A1.  
PN 19-NOV-1998.  
PD 13-MAY-1998; 98WO-AU000352.  
XX 13-MAY-1997; 97AU-00006784.  
PR 26-MAY-1997; 97AU-00008791.  
XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
PA Choo K, Du Sart D, Cancilla MR;  
XX WPI; 1999-009773/01.  
DR New isolated nucleic acid comprising neocentromere sequences from  
XX eukaryotic chromosome - used to produce replicable, segregating  
PT artificial chromosomes that can carry large amounts of DNA for gene  
PT therapy.  
XX Claim 9; Fig 16A; 540bp; English.  
PS The present sequence represents the NC-config derived from a mutated  
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual  
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is  
CC mitotically stable and contains a functional neocentromere at a location  
CC regarded as non-centromeric. This neocentromere maps to q25.2 on  
CC chromosome 10. The specification describes nucleic acid sequences derived  
CC from a eukaryotic chromosome, including a neocentromere or its functional  
CC derivative or hybrid, that are able, in a compatible cell, of  
CC replicating, acting as extra-chromosomal element and segregating during  
CC cell division. The sequences can be used to construct artificial  
CC chromosomes for use in gene therapy comprising a replicable, segregating  
CC nucleic acid that confers a specific phenotype on cells. Human artificial  
CC chromosomes can propagate in human cells and carry large amounts of DNA  
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not  
CC mutagenic. The artificial chromosomes are also useful for generation of  
CC transgenic plants and animals, in production of proteins and to make  
CC diagnostic reagents, e.g. for expression of cytokines, receptors and  
CC growth factors, or to increase the copy number of a gene in a cell. The  
CC constructs may also be used for functional and structural analysis of  
CC chromosomes  
XX  
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;  
Query Match 80.6%; Score 1611.4; DB 2; Length 80240;  
Best Local Similarity 97.4%; Pred. No. 2,4e-264;  
Matches 1666; Conservative 2; Mismatches 40; Indels 3; Gaps 3;

QY 232 GATGATGCTGTCATTAAGATGCTGTATTCATTCCTGATAGATTTCTTTCTAAT 351  
DB 2 GATGATGCTGTCATTAAGATGCTGTATTCATTCCTGATAGATTTCTTTCTAAT 351  
QY 352 ATTTTTCGATGCTGCTGATTAACATACCACTGCTTTTAAATTAAGATTT 411  
DB 62 ATTTTTCGATGCTGCTGATTAACATACCACTGCTTTTAAATTAAGATTT 120  
QY 412 TATGATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTATAC 471

Db 121 TATGGAAAAATTTTCCCTTTTATTTTAAAGTTTATGATGTTATGTTCCCTAATAC 180  
Oy 472 TTAAGTGGGTCTTATAGGAGATATATCTGGGTCTTGATGATATATTAATCTGAT 531  
Db 181 TTAAGTGGGTCTTATAGGAGATATATCTGGGTCTTGATGATATATTAATCTGAT 240  
Oy 532 AATCTCAACCTTTTGTGGAGTGTGGCCATTACATTTAGTGAATTAATAGACATG 591  
Db 241 AATCTCAACCTTTTGTGGAGTGTGGCCATTACATTTAGTGAATTAATAGACATG 300  
Oy 592 GTTGAATTTGCTATACATCTTTTCATTTGTTATATGAGCCATCTTTTCATTTTC 651  
Db 301 GTTGAATTTGCTATACATCTTTTCATTTGTTATATGAGCCATCTTTTCATTTTC 360  
Oy 652 TTTTTCATCTTTGACCATTTTCTTATGATGAAATACCTTTTGTGATTTTCAATATAC 711  
Db 361 TTTTTCATCTTTGACCATTTTCTTATGATGAAATACCTTTTGTGATTTTCAATATAC 420  
Oy 712 TATGGCTTTTATGATATACCTTAAATTTTCTGTTTATGAGATTTATTA 771  
Db 421 TATGGCTTTTATGATATACCTTAAATTTTCTGTTTATGAGATTTATTA 480  
Oy 772 TATACATCTTTAATCTTATACAGATTACCTTCAAAATGATTTTACAGCTCAAGTGA 831  
Db 481 TATACATCTTTAATCTTATACAGATTACCTTCAAAATGATTTTACAGCTCAAGTGA 540  
Oy 832 TGTAGAACCTTACAGAGATATTTTCACTTCTGCTCAATTTTATATGCTA--TGCT 890  
Db 541 TGTAGAACCTTACAGAGATATTTTCACTTCTGCTCAATTTTATATGCTA--TGCT 600  
Oy 891 ATTAATACATTTAGTGTGTTGTTGTTTACCTTATTTGCTGTTGGCTGGGCTCAGCA 950  
Db 601 ATTAATACATTTAGTGTGTTGTTGTTTACCTTATTTGCTGTTGGCTGGGCTCAGCA 660  
Oy 951 AACATTTTCTGTAAAGGCTAGATAGTACAGGCAATCTTGAAGATATCTGAGGTTTGT 1010  
Db 661 AACATTTTCTGTAAAGGCTAGATAGTACAGGCAATCTTGAAGATATCTGAGGTTTGT 720  
Oy 1011 TCCATACCCCAACATTAATACAAATATGGAAGATGATATCACAAATTAAGTGTAC 1070  
Db 721 TCCATACCCCAACATTAATACAAATATGGAAGATGATATCACAAATTAAGTGTAC 780  
Oy 1071 ACAAGCTTTTGGCTTCCAGTGCATATTAAGTTTGTCTTATCTACACTGATCTGT 1130  
Db 781 ACAAGCTTTTGGCTTCCAGTGCATATTAAGTTTGTCTTATCTACACTGATCTGT 840  
Oy 1131 TAAAGTGTGCAATAGTGTATGTCTAATAAAACAACATACCTTAATTTTAAATGCTTAT 1190  
Db 841 TCAAGTGTGCAATAGTGTATGTCTAATAAAACAACATACCTTAATTTTAAATGCTTAT 900  
Oy 1191 ACTAAAAAATGCTAACATCATTTTGACATTCAGTGAATTTTCTTTTGTGCTGTGA 1250  
Db 901 ACTAAAAAATGCTAACATCATTTTGACATTCAGTGAATTTTCTTTTGTGCTGTGA 960  
Oy 1251 AGGCTTTTCTTATGATGATGAGGGGGCTGAGGCTGAGGCTTAAAGGCTGGCTGTG 1310  
Db 961 AGGCTTTTCTTATGATGATGAGGGGGCTGAGGCTGAGGCTTAAAGGCTGGCTGTG 1020  
Oy 1311 CAGTTTCTTAAAAACAAGTGAAGATTGCAATATCACTTCTTCTTCAATGAAGA 1370  
Db 1021 CAGTTTCTTAAAAACAAGTGAAGATTGCAATATCACTTCTTCTTCAATGAAGA 1080  
Oy 1371 TTTTCTCTAGTGTGATGCTTTTGTATGCAATTTTATGACAGATGAACTTTTGA 1430  
Db 1081 TTTTCTCTAGTGTGATGCTTTTGTATGCAATTTTATGACAGATGAACTTTTGA 1140  
Oy 1431 AATTTGA--TCAATCCTCTCAACCCGCTGCTTAAACAACCTAATATTAATAT 1489  
Db 1141 AATTTGAAGTCAATCCTCTCAACCCGCTGCTTAAACAACCTAATATTAATAT 1200  
Oy 1490 CTGAATCATTTGTTGTCATTTCAACAATTTTCAAGTGTCTTCAACAGAGTATGATCCA 1549  
Db 1201 CTGAATCATTTGTTGTCATTTCAACAATTTTCAAGTGTCTTCAACAGAGTATGATCCA 1260

Oy 1550 TCTCATTTTCTGAGATGAAATCTTTGCTCATTCATAAGAAATTCCTCATCTGTCCA 1609  
Db 1261 TCTCATTTTCTGAGATGAAATCTTTGCTCATTCATAAGAAATTCCTCATCTGTCCA 1320  
Oy 1610 GTTATATCATGATGATGAGCAATATACATGCTATGCTTCAAGGCTCATCTTCAAT 1669  
Db 1321 GTTATATCATGATGATGAGCAATATACATGCTATGCTTCAAGGCTCATCTTCAAT 1380  
Oy 1670 CCAGTGTCTGCTGTTTCTACCATCTGCTGTTCTCTCTCATTTGAAGCTTGAAC 1729  
Db 1381 CCAGTGTCTGCTGTTTCTACCATCTGCTGTTCTCTCTCATTTGAAGCTTGAAC 1440  
Oy 1730 TCTCCAAGTTCATCCATGAGGCTGGAATGCACTTCTTCAAAATTCCTGTAATATTA 1789  
Db 1441 TCTCCAAGTTCATCCATGAGGCTGGAATGCACTTCTTCAAAATTCCTGTAATATTA 1500  
Oy 1790 TTTTGAACCTCCCATGATCATGAAATGTTCTTAATAGCACCTGGAATGCTGTAAT 1849  
Db 1501 TTTTGAACCTCCCATGATCATGAAATGTTCTTAATAGCACCTGGAATGCTGTAAT 1560  
Oy 1850 AAAAGCTTTCAATTTACTTATAGTCCAGATCCATCCATCCAGAGATCCACTTCAATGCC 1909  
Db 1561 AAAAGCTTTCAATTTACTTATAGTCCAGATCCATCCATCCAGAGATCCACTTCAATGCC 1620  
Oy 1910 AGTTATACCTTATGGAATGTAATTTCTTCAATATAGGCTTGAAGTTGAATTAATCTCC 1969  
Db 1621 AGTTATACCTTATGGAATGTAATTTCTTCAATATAGGCTTGAAGTTGAATTAATCTCC 1680  
Oy 1970 TTGATCATTTTCTGCAAAATATGATGTTGTG 2000  
Db 1681 TTGATCATTTTCTGCAAAATATGATGTTGTG 1711

RESULT 3  
ADS31115/c  
ID ADS31115 standard; DNA; 2418 BP.  
XX  
XX  
AC ADS31115;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
XX  
DE Human genome high complexity repeat found in the HIRA gene #148.  
XX  
XX  
KW Human; ds;  
KW histone cell cycle regulation defective; S. cerevisiae homologue A; HIRA;  
KW high complexity repeat; in situ hybridisation; Southern blot;  
KW chromosome breakpoint; inherited genetic disease; neoplastic disorder;  
KW chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.  
XX  
OS Homo sapiens.  
XX  
XX  
PN US2003224356-A1.  
XX  
XX  
PD 04-DEC-2003.  
XX  
PF 14-MAY-2001; 2001US-00854867.  
XX  
PR 16-MAY-2000; 2000US-00573080.  
XX  
PA (KNOU/) KNOU J H M.  
XX (ROGA/) ROGAN P K.  
XX  
PI Knoll JHM, Rogan PK;  
XX  
XX  
XX WPI; 2002-062378/08.  
XX  
XX  
XX Single copy genomic hybridization probes for detecting specific nucleic  
XX acid sequences in sample by in situ hybridization useful for detection of  
XX acquired or inherited genetic diseases.  
XX  
XX Example 1; SEQ ID NO 148; 30bp; English.  
XX

CC The invention relates to a nucleic acid hybridisation probe comprising a  
CC labelled, single copy nucleic acid of at least 50 nucleotides, which  
CC will hybridise to a deduced single copy sequence interval in target  
CC nucleic acid (TNA) of known sequence. The single copy sequence is deduced  
CC by comparing the target nucleic acid (e.g. a disease causing gene) with a  
CC collection of high and low complexity repeat sequences as found in the  
CC genome of the organism from containing the target nucleic acid. The probe  
CC is generated by PCR on the target sequence. The probe is essentially free  
CC of blocking nucleic acid sequences which will hybridise to repeat  
CC sequences within the genome of which the TNA is a part, and is labelled  
CC with a label selected from fluorochrome-responsive labels, fluorochromes,  
CC calorimetric chemical, conjugated proteins, antibodies, antigens and  
CC their mixtures. The probe is useful in a hybridisation method, where the  
CC hybridisation method is from in situ hybridisation, Southern blot, and  
CC other methods in which nucleic acid is immobilised, where the method  
CC further comprises selecting a single copy nucleic acid which will  
CC hybridise to a duplicon or triplicon sequence domain. The probe is useful  
CC for determining the existence of previously unknown repeat sequence  
CC families in a genome. The method comprises reacting a labelled probe with  
CC the genome, causing the probe to hybridise and ascertaining if the probe  
CC hybridises to the genome at more than three preferably ten different  
CC locations as a determination of new repeat sequence family, where the  
CC determining step comprises selecting the single copy sequence from a  
CC duplicon or triplicon sequence domain. The probe is useful for  
CC determining a chromosome breakpoint and is useful in the fields for  
CC cytogenetics and molecular genetics for determining the presence of  
CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.  
CC the probes may be used to analyse specific chromosomal locations by in  
CC situ hybridisation as a detection of acquired or inherited genetic  
CC diseases especially for detection of genetic or neoplastic disorder.  
CC Unlike prior art techniques, the probe permits more precise chromosomal  
CC breakpoint determinations by in situ hybridisation. The genomic sequence  
CC comprising the human HIRA gene (histone cell cycle regulation defective,  
CC S. cerevisiae, homologue A) was analysed for single copy sequence  
CC intervals for use as probes of the invention. HIRA is located on  
CC chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and  
CC Velo-Cardio-facial syndromes. The present sequence is a high complexity  
CC repeat found within the human genome used to analyse the HIRA gene for  
CC repeat regions. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from USPTO at [seqdata.uspto.gov/sequence.html?DocID=20030224356](http://seqdata.uspto.gov/sequence.html?DocID=20030224356).

Sequence 2418 BP; 753 A; 472 C; 517 G; 668 T; 0 U; 8 Other;

Query Match 31.8%; Score 636.8; DB 7; Length 2418;  
Best Local Similarity 82.1%; Pred. No. 5; Se-99;

Matches 859; Conservative 1; Mismatches 133; Indels 53; Gaps 9;

```
QY 979 CAGGATACCTTGGAGATCTGTGGTTGGTTCATACCAACCAATATATCAATATAG 1038
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2418 CAGGATACCTTGGAGATCTGTGGTTGGTTCATACCAACCAATATATCAATATAG 2370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1039 CAAGAAGTGATATACATTAAGTGAAGTCAACAAGCTTTGGCTTCCAGTGCATAT 1098
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2369 ----AAGCAATATGCAATTAAGGAGTCAACAATTTTGGTTCCAGTGCATAT 2314
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1099 AAAAGTTTGTCTTACTACTGTAGTGTGTTAAGTGCATTAAGTGTATGTCTTAAA 1158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2313 AAAAGTTTGTCTTACTACTGTAGTGTGTTAAGTGCATTAAGTGTATGTCTTAAA 2254
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1159 AA----ACACATACCTTAATTTTAAATGCTTATTAATAAATGCTAACAATCATTT 1214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2253 AACATGTACATTAATTTTAAATGCTTATTAATAAATGCTAACAATCATTT 2194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1215 GAGCATTCAGTGAAGTGTATCTTTTGTGTTGAAGTCTTTTCTTATGATGACTGA 1274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2193 GAGCCTTCAGGAGGCTAATCTTTTGTGTTGAAGGCTTTGCTTGATGATGATGAC 2134
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1275 T-----CGGGGTGACAGTCTGAAGCTTGAAGTGTGCTGTCGAGTTTCTTAAA- 1322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2133 TGCTGACTGATCAAGGCTGTGTGCTGAAGGTTGGGCTGTGGCAATTTTCTTAAA 2074
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1323 ----ACAACAGTGAAGATTGCAATATCAGTTGACTCTCTTTCATGAAGAATTTCTCTC 1378
```

```
DB 2073 TAAACAAATGAAGTTGGCCGATCAATTAAGCTTCTTCAATGAAGAATTTCTCTC 2014
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1379 TAGTGTGATGCTTTTGTATGCAATTTTATGCAAGTGAATCTTTGAAAATTTGCA- 1437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2013 TAGCATGTGATGCTTTTGTATGCAATTTTATGCAAGTGAATCTTTGAAAATTTGAG 1954
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1438 TCAATCCTCAAAACCCGCTGCTGTTTAAACAACCTAAGTTAATATATCTGAATCC 1497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1953 TCAATCCTCAAAACCCGCTGCTGTTTATCACTAAGTTAATATATCTGAATCC 1894
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1498 ATTTGTTCAATTTCAACAATTTTCAAGTGTCTTCCACAGAGTGAATTCATCATTT 1557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1893 TTTGTTGATTTTCAACATGTTTCAAGATTTTCAACAGAGTGAATTCATCATCA 1834
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1558 CCTGATGGAATCTTTGCTCATCCATGAAGAATAATTCATCTGTCAAGTTTATC 1617
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1833 AACCACTT-----CTTCTCTCATCCATGAAGAATAATTCATCTGTCAAGTTTATC 1778
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1618 ATGAGATGAGCAATATAGATCATGCTTCCAGGCTCACTTCACTTAAATCCAGTTCT 1677
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1777 ATGAGATGAGCAATATAGATCATGCTTCCAGGCTCACTTCACTTAAATCCAGTTCT 1723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1678 CTGCTGTTTTCACACATCTGTGCTCTTCTCTCATTTGAAGCTTGAACCTTCCAA 1737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1722 CTGCTATTTTCCACACATCTGTGCTCTTCTCTCATTTGAAGCTTGAACCTTCCAA 1663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1738 TCATCATGAGGTTGGAATGACTCTTCCAAATCTCTGTTAATATATATATTTGA-- 1795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1662 TCATCATGAGGTTGGAATGACTCTTCCAAATCTCTGTTAATATATATATTTGACC 1603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1796 -CTTCCATGATCATGAATGTTCTTAATGAGCACTGGAATGGAATCTTTCCAAAG 1854
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1602 TCTTCCATGATCATGAATGTTCTTAATGAGCACTGGAATGGAATCTTTCCAAAG 1543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1855 GTTTTCATTTACTAGTCCAGATCCATCCAGAGATCCATTTCAATGACAGTTA 1914
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1542 GTTTTCATTTACTAGTCCAGATCCATCCAGAGATCCATTTCAATGACAGTTA 1488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1915 TAGCCTTAATGATATATTTCTTCAATATATATGAGCTTGAAGTTGAATTAATCTCTGAT 1974
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1487 TAGCCTTAATGATATATTTCTTCAATATATATGAGCTTGAAGTTGAATTAATCTCTGAT 1428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1975 CCATTTTTCGCAAAATATATGTTGTG 2000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1427 CCATGGGCTGCAAGATGATGTGTG 1402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 4
ADY36503/c
ID ADY36503 standard; DNA; 2418 BP.
AC ADY36503;
XX
XX
XX
DT 05-MAY-2005 (first entry)
XX
DB HIRA genomic fragment SBQ ID NO 148.
XX
XX hybridization; DNA detection; neoplasm; genetic disorder; cytogenetics;
XX HIRA; ds.
XX
XX Homo sapiens.
XX
XX WO200188089-A2.
XX
XX 22-NOV-2001.
XX
XX 15-MAY-2001; 2001WO-US015674.
XX
XX 16-MAY-2000; 2000US-00573080.
XX
XX 14-MAY-2001; 2001US-00854867.
XX
XX (CHIL-) CHILDREN'S MERCY HOSPITAL.
```

XX Knoll JHM, Rogan PK, Cazarro PM;  
 XX  
 XX WPI; 2002-062378/08.

XX Single copy genomic hybridization probes for detecting specific nucleic  
 PT acid sequences in sample by in situ hybridization useful for detection of  
 PT acquired or inherited genetic diseases.

XX Example 1; SEQ ID NO 148; 67pp; English.

XX The invention describes a nucleic acid hybridization probe (I) comprising  
 CC a labeled, single copy nucleic acids of at least 50 nucleotides, which  
 CC will hybridize to a deduced single copy sequence interval in target  
 CC nucleic acid (TNA) of known sequence. (I) is useful in a hybridization  
 CC method which comprises preparing a reaction mixture comprising TNA and  
 CC (I) which hybridizes to TNA, and causing (I) to hybridize to TNA, where  
 CC the hybridization method is from in situ hybridization, Southern blot,  
 CC and other methods in which nucleic acid is immobilized, where the method  
 CC further comprises selecting a single copy nucleic acid which will  
 CC hybridize to a duplication or triplicon sequence domain. (I) is useful for:  
 CC determining the existence of previously unknown repeat sequence families  
 CC in a genome; determining a chromosome breakpoint and in the fields of  
 CC cytogenetics and molecular genetics for determining the presence of  
 CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.  
 CC the probes may be used to analyze specific chromosomal locations by in  
 CC situ hybridization as a detection of acquired or inherited genetic  
 CC diseases especially for detection of genetic or neoplastic disorders.  
 CC Unlike prior art techniques, (I) permits more precise chromosomal  
 CC breakpoint determinations by in situ hybridization. Hybridization  
 CC techniques utilizing (I), have made it possible to obtain reliable,  
 CC easily detectable signals with relatively small probes. A reliably  
 CC detectable signal was obtained with a probe on the order of 2 kb in  
 CC length, using fluorescent in situ hybridization (FISH) technology. This  
 CC sensitivity of (I) is improved compared to the prior art, because the  
 CC probes of (I) are homogeneous single copy sequences. However, smaller  
 CC amplified segments, each comprising non-repetitive sequences, may also be  
 CC used in combination as probes to achieve adequate signals for in situ  
 CC hybridization. Complex single copy probes that hybridize to duplicated or  
 CC triplicated targets can also increase hybridization signals. This  
 CC sequence represents a human HIRA genomic sequence that shows homology to  
 CC a known high-complexity repeat sequence family of the human genome and is  
 CC used in the creation of an HIRA gene probe.

XX Sequence 2418 BP; 753 A; 472 C; 517 G; 668 T; 0 U; 8 Other;

XX Query Match 31.8%; Score 636.8; DB 7; Length 2418;

XX Best Local Similarity 82.1%; Pred. No. 5.5e-99;

XX Matches 859; Conservative 1; Mismatches 133; Indels 53; Gaps 9;

QY 979 CAGGATACCTGGAGATCTGTGGTGGTTCCTACACCAATTAATCAATATG 1038  
 DB |||||  
 DB 2418 CAGGATACCTGGAGATATGCGGGTTCGACACCGCAAT- 2370  
 QY 1039 CAAGAAGTATATCAATTAAGTACACCAAGCTTTTGGCTCCAGTGCATAT 1098  
 DB |||||  
 DB 2369 -AAGCAATATGCAATTAAGCAGTCAACGATTTTGGTTCCAGTGCATAT 2314  
 QY 1099 AAAAGTTTGTCTTAATCACTAGTACTGTTAAAGTGCATAGTGTATGCTTAA 1158  
 DB |||||  
 DB 2313 AAAAGTTTGTCTTAATCACTAGTACTGTTAAAGTGCATAGTGTATGCTTAA 2254  
 QY 1159 AA-ACGATACCTTAATTTAAATGCTTATTAATTAATTAATTAATTAAT 1214  
 DB |||||  
 DB 2253 AACATGTATCACTTAATTTAAATGCTTATTAATTAATTAATTAATTAAT 2194  
 QY 1215 GAGCATTCAGTGTATGCTTTTGGCTGTGGAAGGCTTTTCTTATGATGACTGA 1274  
 DB |||||  
 DB 2193 GAGCTTCAGGAGTGTATGCTTTTGGCTGTGGAAGGCTTTTCTTATGATGAG 2134  
 QY 1275 T-----CGGGGTTCAGGTGCTGAAGCTTAAAGGTGCTGTGGCACTTTCTTAA- 1322  
 DB |||||  
 DB 2133 TGCTGATGATCAAGGTGGTGGTGTCTGAAGGTTGGGGTGGCTGTGGCAATTTCTTAA 2074

QY 1323 ----ACAAAGTGAAGATTCGAATATCAGTTGACTCTTCTTTCATGAAAGATTTCTCTC 1378  
 DB |||||  
 DB 2073 TAAGACAACAATGAAGTTGGCCGATCAATGACTCTTTCATGAAAGATTTCTCTG 2014  
 QY 1379 TAGTGTGTATGCTTTTGTATAGATTTTATGCAACAGTAAAGCTTTTGAAGATGGA- 1437  
 DB |||||  
 DB 2013 TAGATGTGATGCTGTTTGAAGATTTTATCCACAGTAAAGCTTTTCAAAATGGAG 1954  
 QY 1438 TCATTCCTCAACCCGTCTGCTTAAACAACTTAAGTTAATATATTCGAATCC 1497  
 DB |||||  
 DB 1953 TCATTCCTCAACCCGTCTGCTTAAACAACTTAAGTTAATATATTCGAATCC 1894  
 QY 1498 ATTGTGTATTCATCAACATTTTCAAGTGTCTTACCAAGATGATTCATTCATTT 1557  
 DB |||||  
 DB 1893 TTTGTGTATTCATCAACATGTTCAAGCATCTTACAGAGATGATTCATTCACAGA 1834  
 QY 1558 CCTGATGGAATCTTTGCTCATCAATGAAGAAATCCATCTGTTCAAGTTTATC 1617  
 DB |||||  
 DB 1833 AACCACTT-----CTTGTCTATCATTAAGAAAGCACTCTCATCCGTTCAAGTTTATC 1778  
 QY 1618 ATGAGATTCAGCAATACAGTCAATGCTTCAGGCTCACTTCACTTAAATTCAGTTCT 1677  
 DB |||||  
 DB 1777 ATGAGATTCAGCAATACAGTCAATGCTTCAGGCTCACTTCACTTAAATTCAGTTCT 1723  
 QY 1678 CTGCTGTTTCTACCAATCTGTGCTTCTCTTCATTTGAAGTCTTCAACCTCTCCAG 1737  
 DB |||||  
 DB 1722 CTGCTGTTTCTACCAATCTGTGCTTCTCTTCATTTGAAGTCTTCAACCTCTCCAG 1663  
 QY 1738 TCATTCAGAGGTTGGAATGAGATCTTCCAAATTCGTTAATATTAATTTTGA-- 1795  
 DB |||||  
 DB 1662 TCATTCAGAGGTTGGAATGAGATCTTCCAAATTCGTTAATATTAATTTTGAAC 1603  
 QY 1796 -CTCCATGATCATGATGTTCTTAATGAGCACTGGAATGATGATCTTTCCAAAG 1854  
 DB |||||  
 DB 1602 TCTTCATGATCATGATGTTCTTAATGAGCACTGGAATGATGATCTTTCCAAAG 1543  
 QY 1855 GTTTCAATTTAATTTAGTCCAGATTCATTCAGAGATTCATTCATTCATTCAGTTA 1914  
 DB |||||  
 DB 1542 GTTTCAATTTAATTTAGTCCAGATTCATTCAGAGATTCATTCATTCATTCAGTTA 1488  
 QY 1915 TAGCTTATGATGATTTTCTCAATTAATTAAGGCTTGAAGTTGAATTTCTCTTAT 1974  
 DB |||||  
 DB 1487 TAGCTTATTAATTTTCTTAATTAATTAAGGCTTGAAGTTGAATTTCTCTTAT 1428  
 QY 1975 CCATTTCTGCAAAATAGATGTTG 2000  
 DB |||||  
 DB 1427 CCATGGCTGCAAGATGATGTTG 1402

RESULT 5  
 ADH22218  
 ID ADH22218 standard; DNA; 14033 BP.  
 XX  
 AC ADH22218;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Human cholinergic receptor, nicotinic, gamma (CHRN) genomic DNA SegID 1.  
 XX  
 XX human; ds; chromosome 2q33-q34; cholinergic receptor, nicotinic, gamma;  
 KW CHRN; haplotype; drug discovery; acetylcholine receptor; AChR;  
 KW myasthenia gravis; screening assay; single nucleotide polymorphism; SNP;  
 KW gene.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT variation replace(3697,t)  
 FT /tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3883,t)  
 FT /tag= b

```
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3902,g)
FT /tag= c
FT /standard_name= "Single nucleotide polymorphism"
FT 4066. .10034
FT /tag= e
FT /product= "CHRNA protein"
FT /note= "This coding sequence contains 12 exons"
FT 4066. .4120
FT /tag= d
FT /number= 1
FT 4121. .4309
FT /tag= f
FT /number= 1
FT replace(4198,c)
FT /tag= g
FT /standard_name= "Single nucleotide polymorphism"
FT 4310. .4449
FT /tag= h
FT /number= 2
FT 4450. .4698
FT /tag= i
FT /number= 2
FT 4699. .4743
FT /tag= j
FT /number= 3
FT 4744. .4919
FT /tag= k
FT /number= 3
FT 4920. .5029
FT /tag= l
FT /number= 4
FT 5030. .5691
FT /tag= m
FT /number= 4
FT 5692. .5847
FT /tag= n
FT /number= 5
FT 5848. .6740
FT /tag= o
FT /number= 5
FT replace(6606,g)
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6728,t)
FT /tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT 6741. .6838
FT /tag= r
FT /number= 6
FT 6839. .7199
FT /tag= s
FT /number= 6
FT replace(7124,c)
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT 7200. .7400
FT /tag= u
FT /number= 7
FT replace(7257,g)
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT replace(7258,a)
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT 7401. .7592
FT /tag= x
FT /number= 7
FT 7593. .7707
FT /tag= y
FT /number= 8
FT 7708. .7902
FT /tag= z

FT /number= 8
FT replace(7847,g)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT replace(7891,t)
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT 7903. .8017
FT /tag= ac
FT /number= 9
FT 8018. .8684
FT /tag= ad
FT /number= 9
FT replace(8042,c)
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT replace(8621,a)
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism"
FT 8685. .8898
FT /tag= ag
FT /number= 10
FT 8899. .9089
FT /tag= ah
FT /number= 10
FT 9090. .9220
FT /tag= ai
FT /number= 11
FT 9221. .9860
FT /tag= aj
FT /number= 11
FT replace(9337,a)
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT 9861. .10034
FT /tag= al
FT /number= 12
FT replace(9902,t)
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism"
FT replace(10111,c)
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism"

PN WO200222643-A1.
PD 21-MAR-2002.
PP 17-SEP-2001; 2001WO-US029206.
PR 15-SEP-2000; 2000US-0232807P.
PA (GENA-) GENAISSANCE PHARM INC.
PI Gilson CR, Koshy B, Kiem SE, Sauker EA;
XX WPI; 2002-371968/40.
XX P-PSDB; ADH22134.
XX
PT New genetic variants of cholinergic receptor, nicotinic, gamma
PT polypeptide, CHRNA gene useful for therapeutic purposes and for
PT expressing CHRNA protein useful in identifying drugs to treat myasthenia
PT gravis.
XX
PS Claim 1; SEQ ID NO 1; 107pp; English.
XX
CC This invention relates to novel genetic markers and variants of the gene
CC encoding the cholinergic receptor, nicotinic, gamma polypeptide (CHRNA),
CC located on chromosome 2q33-p34. Specifically, it refers to a set of
CC haplotypes in the CHRNA gene, which are useful for improving the
CC efficiency and output of the drug discovery process by the identification
CC of drugs that can target the CHRNA protein and treat disorders associated
CC with its abnormal expression or function. The CHRNA protein is the gamma
```

CC subunit of the acetylcholine receptor (AChR), and autoantibodies directed  
CC against the embryonic form of AChR play an important role in the  
CC pathogenesis of neonatal myasthenia gravis. As such, the present  
CC invention describes a method for identifying an association between a  
CC trait (such as a clinical response to a drug that targets CHRNA) and a  
CC haplotype or haplotype pair of the CHRNA gene. Furthermore, it is useful  
CC in screening assays, for the development of diagnostic tests and for  
CC therapeutic treatments of myasthenia gravis using gene therapy. This  
CC polynucleotide is the human CHRNA genomic DNA sequence that contains the  
CC SNPs of the invention.

XX Sequence 14033 BP; 3341 A; 3737 C; 3747 G; 3192 T; 0 U; 16 Other:

Query Match 26.3%; Score 526.6; DB 6; Length 14033;

Best Local Similarity 78.7%; Pred. No. 3e-80; Mismatches 169; Indels 39; Gaps 10;

DB 1051 ATCCATTAATGAAGTGTGACCA-CAAGCTCTTTGGCTTCCGAGTCATATAAGTTTGC 1109  
DB 12263 AACACAGAGAGCAAGTCAATGTTTGGTTCCTCAATGATATATAAGTTTGT 12322  
QY 1110 TTATCTACTAGTGTGCTGTAAAGTGTGCAATAG--TGTATGTCTTAAATAACAT 1166  
DB 12323 TTACCTTACTGTGACCAAGAGAGTCAATAGCATTTCTAATATAACATATACAT 12382  
QY 1167 ACCCTTAATTTTAAATGCTTTATTAATAAATAATGCTAACATC-ATTGACCATTCAGT 1225  
DB 12383 ACCTAATTTTAAATAATCTTTATGTGCTAAATAATGCTAATATGATGACCTTCGGT 12442  
QY 1226 GAGTGTATCTTTTGTGCTGTGAGAGCTTTTCTTATGATGACTAT----- 1275  
DB 12443 GAGTGTCTCTTTTGTGCTGTGAGAGGCTTGTCCGAGGTTGATGTTGCTGACTGGT 12502  
QY 1276 -CGGGGGTCAGGTGTGAAAGCTTAGGGTGTGAGCTGTGCACTTTCTTAA-----ACAAG 1329  
DB 12503 CAGGGTGTGCTGTGAGAGGTTGGGGTGTGCTGTGCACTTTCTTAAATAAGACATGCG 12562  
QY 1330 TGAAGATGCAATATATCATGTTGACTCTCTTTCATGAAAGATTTCTCTAGTGTGTAT 1389  
DB 12563 TGAAGCTTTTGGCATGAGTGTGACTCTCTTTCATGAAAGATTTCTCTAGACATGCGAT 12622  
QY 1390 GCTTTTGTATGACATTTTATGACAGATGACATCTTCTTGAATAATGGA-TCAATCTCTC 1448  
DB 12623 GCTTTTGTATGACATTTTGTGCAAGTGTGAGCTTCTTCAAAATGAGATCAATCTCTC 12682  
QY 1449 AAACCTGTGCTGCTTAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1508  
DB 12683 AAACCTGTGCTGCTTAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12742  
QY 1509 TTCAACAATTTTCAAGTGTCTTCAACAAGATGATGATTCATCTCAATTTCTGAGATGGA 1568  
DB 12743 TTCAACAATTTTCAAGTGTCTTCAACAAGATGATGATTCATCTCAATTTCTGAGATGGA 12799  
QY 1568 ATCTTTGCTCATTCATTAAGAGAAATTTCTCATCTTGTCAAGTTTATCATGATGATGGA 1628  
DB 12800 -TCTTTGCCATTCATTAAGAGAAATTTCTCATCTTGTCAAGTTTATCATGATGATGGA 12858  
QY 1629 GGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1688  
DB 12859 GGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12913  
QY 1689 TACCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1748  
DB 12914 CACCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12973  
QY 1749 GGTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1805  
DB 12974 GGTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13033  
QY 1806 ATCAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1865  
DB 13034 ATCAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13093

QY 1866 ACTTATGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1925  
DB 13094 GCTTGTCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13148  
QY 1926 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1985  
DB 13149 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13208  
QY 1986 AAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2000  
DB 13209 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13223

RESULT 6

ADM01834  
ID ADM01834 standard; cDNA; 3895 BP.

XX ADM01834;  
XX 20-MAY-2004 (first entry)

DE Human cDNA of the invention SEQ ID NO:519.

XX ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.

XX Homo sapiens.

XX EPI347046-A1.

PD 24-SEP-2003.

PF 12-APR-2002; 2002EP-00008400.

PR 22-MAR-2002; 2002JP-00137785.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,

PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuho Y;

DR WPI: 2003-723558/69.

DR P-PSDB; ADM04277.

PT New polynucleotides and polypeptides are useful in gene therapy, for

PT developing a diagnostic marker or medicines for regulating their

PT expression and activity, or as a target of gene therapy.

PS Claim 1; SEQ ID NO 519; 305bp; English.

XX The invention relates to a novel human polynucleotide and the encoded

XX polypeptide. A polynucleotide of the invention may have a use in gene

XX therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful

XX as a primer for synthesizing the polynucleotide or as a probe for

XX detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are

XX useful in gene therapy, for developing a diagnostic marker or medicines

XX for regulating their expression and activity, or as a target of gene

XX therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides

XX are useful as pharmaceutical agents. The present sequence represents a

XX cDNA sequence of the invention.

XX Sequence 3895 BP; 1139 A; 804 C; 807 G; 1145 T; 0 U; 0 Other;

Query Match 25.8%; Score 516; DB 11; Length 3895;

Best Local Similarity 77.5%; Pred. No. 1.8e-78; Mismatches 175; Indels 40; Gaps 8;

DB 1071 ACAAGCTTTTGGCTTCCAGAGTCAATATAAGTTTGTCTTACTACTAGTGTCTGT 1130

DB 1362 ATTAATTTTGTGTTTCCAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1421

QY 1131 TAAAGTGTCAATAGTGTGTATGCTTAAATAA-----CACATACCTTAATTTTAAATGCG 1184

Db	1422	TAAGGGTCATATGACATTATGCTGAAAAAACAATGCTCANTGCCCTTAATTTAAAAATAC	1481
OY	1185	TTTATATTAAAAATGCTAACATCATTTGAGCATTCAGAGAGTTGTAATCTTTTGCT	1244
Db	1482	TTTATATTAAAAATGCTAGCAATCATCTAACCTTCAGAGATCATAAATATTTTGCT	1541
OY	1245	GGTGAAGATCTTTTCTTATTGATGACTGAT-----CGGGGTCAGGGTCGAA	1293
Db	1542	AGTGAAGGCCCTTGCTGCTGGATGTTGATGATGGCTGCTGACATCAGATGATGATGTTGCTGAA	1601
OY	1294	GCTTAGGGTGGCTGTGCGAGTTTCT---TAAACAACAGTGAAGATTGCAATATACGT	1348
Db	1602	GGTTAGGGTGGCTGGCGCAATTTCTTAACAATAGCAACAGTGAAGTTCTGTGATTCGAT	1661
OY	1349	TGACCTCTCTTTCATGGAAGATTCTCTAGTGTGATGATGATGATCTTTTGAATAGATTTTA	1408
Db	1662	TGACCTCTCTTTCATGGAATATTTCTTATAGCATGTAAAGTGGTTGATGATGATTTTA	1721
OY	1409	TGCACATAGAACCTTCTTGGAAAAATGGA-TCAATCCCTGCATAACCCGCTGCTCTTAA	1467
Db	1722	CCCAAGATAGATTTATTTACGAATAGAAATGAAATGACGCTCTAAACCGTGCATCTTTTA	1781
OY	1468	CACTTAAGTTAATTAATTAATTTGATTCATTTGTCATTTCAACAATTTTCAACGtG	1527
Db	1782	TTAATTAAGTTACAGCATATTTCTAAACCTTTGTCATTTCAACAATTTTCAACA	1841
OY	1528	TCTTCACAGAGATGATTTCCATCTCATTTTCTGTGATGAGAAATCTTTGCTCATCATAG	1587
Db	1842	TCTTCACAGAGATGATTTCTGTCTCTCAAGAAACCACTT---TCTTTGTCATTTGATAG	1897
OY	1588	AAGAATTTCTCATCTGTTCAGATTTTATCATGATGATGAGCAATACATCATGCTGTC	1647
Db	1898	AAGCAACCCCTCATCTGTTCAGATTTTCTCATGATGATGAGCAATACATCATTTTC	1957
OY	1648	AGGCTCACCTCATCTTTAATTTCCAGTCTCTGTGCTGTTCTTACACATCTGTGCTCT	1707
Db	1958	AGG-----CTCATTTTAAATTCATGATCTCTTGCTATTTTACACATCTGCAGTTACT	2012
OY	1708	TCTTCATTTGAAGTCTTGAACCTCTCCAAATCATCATGAGGTTGGAATGCACTTCTTC	1767
Db	2013	TCTTCATTTGAAGGCTTGAACCCCTCAAAATGACCAATGAGGGTTGGAATCACTTCTTC	2072
OY	1768	CAAAATCTCTGTAATATTATTAATTTTGAACCTCC--CATGAATCATGAATGTTCTTAATG	1824
Db	2073	CAAAATCTCTGTAATGTTGATGTTTGAACCTCCCTCATGATGATGATAATGATTTTATG	2132
OY	1825	GCACCTGGAATGATGAATCTTTCGAAAAGGTTTTCATTAATTAAGTCCAGATCCATCC	1884
Db	2133	ACAATCTAAATATGATCAGTCTTTCCAGAAAGTTTCAATTTACTTTGCCCAATCCAT--	2190
OY	1885	ATCCAGAGATTCACCTTTCATGSCCAGTTATAGCCTTAATGATGATGATTTCTTCATATAT	1944
Db	2191	---CAGAGGATCACTGCTATGCAACTGATGAGCTTAATAAAGTATTTCTTAATAATAT	2247
OY	1945	AAGGCTTGAAGTTGAATTAATCTCTGATGATCATTTTCTGCAAAATGATGTTGTG	2000
Db	2248	ATGCTTGAAGTCTAAATGATCTCTTGATTCATTAAGTGTGAGATGAGGAGTGTGTG	2303

RESULT 7  
ABD32923\_7  
Continuation (8 of 8) of ABD32923 from base 700001 (Human cancer-associated genomic DNA  
and  
NP Sequence split into 8 fragments LOCUS ABD32923 Accession Abd32923

WP	Fragment Name	Begin	End
WP	ABD32923_0	1	11000
WP	ABD32923_1	100001	21000
WP	ABD32923_2	200001	31000
WP	ABD32923_3	300001	41000
WP	ABD32923_4	400001	51000
WP	ABD32923_5	500001	61000
WP	ABD32923_6	600001	71000
WP	ABD32923_7	700001	788759

Query Match	25 3#:	Score 505 2;	DB 13;	Length 88759;
Best Local Similarity	75.5#:	Pred. No. 1,4e-76;		
Matches 743;	Conservative 0;	Mismatch 208;	Indels 33;	Gaps 8;
QY	1034	ATATGCAAGAGTGGATATCATCATTAAGTACGACACAAAGCTCTTTGGCTCCAGTG	1093	
DB	4541	ATTCATTAAGTGAATATTTGGCATTAATAAGTCGACGAATTTTGGTTTCCAGTG	4600	
QY	1094	CATATAAAGTTTGGCTTATCTACACTGTAAGTCTGTTAAGTGTGAATAGTGTATGTC	1153	
DB	4601	CATATAAAGTTATGTTATATATATACGTAGTCTATTTAAGTATCAATAGCATTAATGTC	4660	
QY	1154	TAAAAAAACATACCTTAATTTTAAAAAGCTTATTACTTAAAAAAGCTTAACATCAATT	1213	
DB	4661	TAAAAAATATACATACCTTATTTTAAAGTACCTTATAGCTCAAAAGTACAAACCCACC	4720	
QY	1214	TGAGCATTCAGTGAAGTTGTAATCTTTTGGCGGGAAGGCT-----TTCTTAT	1264	
DB	4721	TGAGCTTCAGTGAAGTTGTAATCTTTTGGCGGGAAGGCTTTCCCAAGTTGAAGTC	4780	
QY	1265	TGATGACTGATCGGAGGCTCAGTGCTGACGCTGAAGCTTGAAGCTGCGAGTCTTTAA--	1322	
DB	4781	TCCTGACTGATCAGAGTGTGATTTCTTAAGTGAAGTATGCTGTGAGAGTTCTTAAT	4840	
QY	1323	--AACAAGTGAAGATTGCATATACAGTTGACTCTTCTTCATGAAGAATTCCTCT	1379	
DB	4841	GAGCAACAGGAAGCTTGCACATCAATTGACTCTTCTTCATGAAGAATTCCTCTGT	4900	
QY	1380	AGTGTGATGCTTTTGTATAGCTTTTATGCAACAGTGAACCTTCTTGAATAATGGA-T	1438	
DB	4901	AGATGTGATGCTTTTGTATAGCTTTTATGCAACAGTGAACCTTCTGTAACGTTGAGT	4960	
QY	1439	CAATCTCTCAAAACCTGCTGCTGTTTAAACAACCTAAGTTAATATAATTTGTAATCCA	1498	
DB	4961	CACACTTCTAAACCTTGCACACTTCTTCACTCACTAATATGTAAGAAATTTCTTAACCTT	5020	
QY	1499	TTGTGTCAATTCACAATTTTCAACAGTGTCTTCCACGAGTGAATTCATCTCATTTTC	1558	
DB	5021	TTGTGTCAATTCACAATGTGTCAATGTAATTTTCCACGAGTGAATTCATCTCAAGAA	5080	
QY	1559	CTGAGATGAATCTTTGCTCATCCATAAGAAATTCCTCATCTGTCTCAAGTTTATCA	1618	
DB	5081	GCCACTT-----TCTTCTCATCTGTAAGAAACAAATGTTCAACCATTTAAGTGCATCA	5136	
QY	1619	TGAGATTGACAGCAATTAAGTCAATGCTCTTCAGGCTCACTCACTTAAATTCAGATCTTC	1678	
DB	5137	TGCAATTTGACAGCAATTAAGTCTCATCTTCAGG-----CTCCCTTTCTAATTTCAATGTTCTC	5191	
QY	1679	TTGCTGTTTCTACCACTCTGTGTGTTCTTCTCTTCATTGAAGCTTGAACCTCTCCAACT	1738	
DB	5192	TTGCTATTTTCTACCACTATGAGATTAAGTACTCTCTCACTGAAGCTTGAACCTCTCAAACT	5251	
QY	1739	CATCCATGAGGCTTGAATGCACTTCTTCCAATTCCTGTTAATATTATATTGACCT	1798	
DB	5252	CATCCATGAGGCTTGAATCAATTTCTTGGCAACTCTGTTTATGTTGATATTGACCT	5311	
QY	1799	CC---CATGAATCAATGATGTTCTTATATGCACTGGAATGATGATTCCTTCCAAAGG	1855	
DB	5312	CTCTCATGAATTAACAAGTATTCGAATGACCTTAAGTGAATTCCTTTCAGAAGG	5371	
QY	1856	TTTTCAATTTACTAGTCCAGATCCATCCATCCAGAGATCCACTTCAATGCGCAATTAT	1915	
DB	5372	TTTTCAATTTACTTGTGCCAGATCCAT-----CAGAAGATCACTATCTATGCGCAGTAC	5425	
QY	1916	AGCCTTAATGAATGATATTTCTTCATATATGAAGCTTGAAGTTGAATTAATCTCTTGATC	1975	
DB	5426	AGCCTTAATGAAGTATATTTCTTAAATGATTAACCTGAAGTCAAAATATATACCTTGATC	5485	
QY	1976	CATTTCTGCAAAATAGATGTGT	1999	
DB	5486	CATGGGCTGCAAGATGATGTGTGT	5509	



```
RESULT 8
ABD32923_6
Continuation (7 of 8) of ABD32923 from base 600001 (Human cancer-associated genomic DNA
WP Sequence split into 8 fragments LOCUS ABD32923 Accession ABD32923
WP Fragment Name Begin End
WP ABD32923_0 1 110000
WP ABD32923_1 100001 210000
WP ABD32923_2 200001 310000
WP ABD32923_3 300001 410000
WP ABD32923_4 400001 510000
WP ABD32923_5 500001 610000
WP ABD32923_6 600001 710000
WP ABD32923_7 700001 788759

Query Match 25.3%; Score 505.2; DB 13; Length 110000;
Best Local Similarity 75.5%; Pred. No. 1.4e-76;
Matches 743; Conservative 0; Mismatches 208; Indels 33; Gaps 8;

QY 1034 ATATGCAAGAGTGGATATCATCAATTAAGTGCACAGAGTCTTTGGCTTCCAGTG 1093
DB 104541 ATCTCAATTAAGTGAATATGCAATTAAGTCCAGAGATTTTGTGTTCCAGTG 104600
QY 1094 CATATTAAGTGGTTGTATATCTACACGTGAGTCTGTAGTGTGCAATAGTATATGTC 1153
DB 104601 CATATTAAGTGTATGTTATATATATCTAGTCTATTAAGTATTAAGATGCAATATATGTC 104660
QY 1154 TAAAAAACAACATACCTTAATTTTAAATGCTTATTAACAAAAATGCTAACATCATT 1213
DB 104661 TAAAAAATATATACATCTTTAATTTTAAAGTACCTTATGCTCAAAAGTACTAACCCCAACC 104720
QY 1214 TGAGCATTCAGTGAAGTGTAAATCTTTTGTGTTGAGAGTCT-----TTTCTTAT 1264
DB 104721 TGAGCCTTCAGTGAAGTGTAAATCTTTTGTGTTGAGAGTGTGCTCAAAAGTGAATGTC 104780
QY 1265 TGATAGCTGATGGGGGTGAGGTGCTGAGGCTTAAGGGTGGTGGCACTTTCTTAA-- 1322
DB 104781 TCTGATCTGATGAGATGTGATTTGTAAAGTAGAGTGGTGGCACTTTCTTAAAT 104840
QY 1323 ---ACACAGTGAAGATGCAATATCATGATGACTCTTCCCTTCATGAAGATTTCTCT 1379
DB 104841 GAGACACAGTGAAGATTTGCCATCATGATGACTCTTCTTTTCATGAAGATTTCTCT 104900
QY 1380 AGTGTGTAGTCTTTTGTATGATTTTATGACAGTGAAGACTCTTTGAAGATTTGA--T 1438
DB 104901 ACCAGTGTAGTGTGTTGTGATCATTTTACTCAGATGAAGACTCTGTGAAGCTTGAAGT 104960
QY 1439 CAATCTCTCCAAACCTGCTGCTTTAACACCTTAAGTAAATATATTCGTAATCCA 1498
DB 104961 CACACTTCTCAACCTTGGCACTTCTTCATCACTATATGATGAATATTTCTAAACCTT 105020
QY 1499 TGTGTGTCAATTTCAACAATTTTCAACAGTGTCTTACACAGAGTGAATTCATCTCATTT 1558
DB 105021 TGTGTGTCAATTTCAACAATTTTCAACAGTGTCTTACACAGAGTGAATTTCTATCTCAAGA 105080
QY 1559 CTGAGATGAATCTTGTCTCATCAAGAGAAATTCCTCATCTGTCAAGTTTATCA 1618
DB 105081 GCCACTT-----TCTTGTCTCATCTGAAGAAACAATGTTCACCACTTTAAAGTTCATCA 105136
QY 1619 TGAGATGAGCAATACAGTCAATGTCTTACAGGCTCACTTCACTTTAATTCAGATTTTC 1678
DB 105137 TGACATTTGAGCAATTCAGTCTCATCTTCAAG-----CTCCCTTTCTAATTTCAAGTTCTC 105191
QY 1679 TTTGCTTTCTTACACATCTGTGTTCCCTTCCATTTGAAGCTTGAAGCTTCCCAAGT 1738
DB 105192 TTGCTATTTCTACACATGAGTGAATGCTTCTCTCACTGAAGCTTGAAGCTTCCCAAGT 105251
QY 1739 CATCCATGAGGTTGGAATGCACTTCTCCAAATCTCTGTTAATTTAATTTTGAACCT 1798
DB 105252 CATCCATGAGGTTGGAATGCAATTTCTTGCCACTCTGTTAATTTGAATTTTGAACCT 105311
QY 1799 CC---CATGATATGAAATGTTCTTAAATGACCTGGAATGTGAATCTTTCCAAAAG 1855
DB 105312 CCTCATGAAATTAACAAGTATCTGAATGGCACTGAAGTGTGAATCTTTTTCAGAAAG 105371
```

```
QY 1856 TTTTCAATTAAGTATGTCACAGATTCATCCAGAGATTCACCTTTCAATGCAAGTTAT 1915
DB 105372 -TTTCAATTAAGTATGTCACAGATTCAT-----CAGAAATCACTATCTATGCGAGCTAC 105425
QY 1916 AGCCTTAAGTAAATGTTCTTCAATTAATAGGCTTGAAGTGAATTAATCTCTGATTC 1975
DB 105426 AGCCTTAAGTAAATGTTCTTCAATTAATAGGCTTGAAGTGAATTAATTAATCTTGAATC 105485
QY 1976 CATTTTGTCAAAATATGATGTTGT 1999
DB 105486 CATGGGCTGCAAGATGATGTTGT 105509

RESULT 9
ABD3512/c
ID ABD3512 standard; DNA; 55829 BP.
XX
XX ABD3512;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human cancer-associated (CA) gene HD07-100.
XX
XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX ds; cancer; cytostatic.
XX
XX Homo sapiens.
XX
XX WO2004058146-A2.
XX
XX 15-JUL-2004.
XX
XX 15-DEC-2003; 2003WO-US040081.
XX
XX 17-DEC-2002; 2002US-00322281.
XX
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX
XX WPI; 2004-499109/47.
XX
XX Novel human cancer associated protein encoded within open reading frame
XX of cancer associated gene, useful as targets for diagnosing cancer.
XX
XX
XX Claim 16; SEQ ID NO 688; 182pp; English.
XX
XX The invention relates to cancer-associated proteins (CAP) and the cancer-
XX associated (CA) nucleic acids encoding them. The invention also relates
XX to a method for treating cancers involving administering to a patient an
XX inhibitor of CAP, and a method of screening for anticancer activity in a
XX potential drug involving providing a cell that expresses a CA gene,
XX contacting a tissue sample derived from a cancer cell with an anticancer
XX drug candidate and monitoring the effect of the anticancer drug candidate
XX on expression of the CA gene. The CAP proteins are useful for detecting
XX cancer associated with a bioactive agent capable of modulating the
XX and for screening for a bioactive agent capable of modulating the
XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing
XX cancer, involving determining the expression of a CA nucleic acid in a
XX tissue. This sequence represents a human CA gene of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 55829 BP; 17201 A; 9903 C; 10825 G; 17900 T; 0 U; 0 Other;
XX
Query Match 25.0%; Score 500.6; DB 13; Length 55829;
Best Local Similarity 69.4%; Pred. No. 8.1e-76;
Matches 857; Conservative 0; Mismatches 329; Indels 49; Gaps 11;

QY 784 ACTTATACAGATTAACCTTCAATATGATATTTTACCAAGCTCAAGTGAATGTGAACCTT 843
DB 105312
```

Db 21679 ACCTGAAAAACATCTATGAAATATAGCTCAGCACTTAACCTGTGACGCAATTC 21620  
Qy 844 ACAAGATATATTTTCATTTCTGTCTCCATTTTATGCTATGTCTATTAATCAATAGG 903  
Db 21619 TAAGCTCTTGGCTCAGAGCCCTTTAAATTTCTTAAATTTACTGCTCTAAATTTCTAAA 21560  
Qy 904 TTTGT 963  
Db 21559 TTAGGGCATATACATGATCAG--TGCCTATTTTGAAGAATTGTGTGTGTGTGTGTGT 21502  
Qy 964 AAGGCTAGATAGTACAGGATACCTTGGAGATCTGTGGGTTTGGTTTCATACACAC 1023  
Db 21501 ATAGGACAGAAATATACAGTATACATGAAATATGCAAGTTTGGGTCAAGATTAACAC 21442  
Qy 1024 AATATATCAAAATATGCAAGAGTATATCAATTAAGTATGACACAAAGCTTTTGG 1083  
Db 21441 AATATATGTAATTTCAAAATATATGCCAGCACATTAATTTT-----TTTTGG 21396  
Qy 1084 CTTCAGTGCATATTAAGTTTGTCTTATCTACACTGTAGTCTGTATTAAGTGCATTA 1143  
Db 21395 TTTCTTAGTGCATATTAAGTTTGTCTTATCTAGTCTGTATTAAGTGTATAC 21336  
Qy 1144 GTCTTATGTCTA-----AAAAACAATACCTTAATTTTAAATGCTTTATTAATAAAAAATGC 1202  
Db 21335 GCATTTATGTGCAGAAAAATGTGATACCTTGTATTTAAACAATTTATTTGCTACAAATGC 21276  
Qy 1203 TTAACATCACTTTAGAGATTCAGAGTGTATCTTTTGTCTGTGTGTGTGTGTGTGTGT 1262  
Db 21275 CAACAATCACTTACAGCTTCAGTGAATTTAATCTTTTGTGTGTGTGTGTGTGTGTGT 21217  
Qy 1263 ATTTGATGATGAT-----CGGGGTGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1311  
Db 21216 CATTTGATGAGCTGCTACATGATCAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 21157  
Qy 1312 AGTTTCTTAAA--ACAACAGTGAAGATTCATATATCATGATGATCTTCTTCATGAAAG 1369  
Db 21156 AATTTCTTAAATTAACATGAAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 21097  
Qy 1370 ATTTCTCTCAGT 1429  
Db 21096 ATTTCTCTGTAGATCAACATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 21037  
Qy 1430 AAATTTGGA--TCAATCTCTCAAAACCTGTCTGTCTTAAACAACCTTAATTAATAATAT 1488  
Db 21036 AAATGAGGACATCTTCTCAAAACCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 20977  
Qy 1489 TCTGAATCATTTGT 1548  
Db 20976 TCTAAATCTTTGT 20917  
Qy 1549 ATCTCATTTCTGTAGATGGAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1608  
Db 20916 ATCTCATGAAACCACTT---TCTTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 20861  
Qy 1609 AGTTTATCATGATTTGAGCAATATCAGTATGCTTCAAGGCTCACTTCACTTTTAAT 1668  
Db 20860 AGTTTATCTTAAGATTTGAATATCTTCAATTTTCAAG-----CTCCACTTTTAAT 20806  
Qy 1669 TCCAGTCTCTGT 1728  
Db 20805 TCAAGTCTCTGT 20746  
Qy 1729 CTCTCAATCATTCATGAGGTTGGAATTCGACTTTCGAAATTCCTGTTAATTAATTA 1788  
Db 20745 CCTCAAAATCATTCAGAGGTTGGAATTCGACTTTCGAAATTCCTGTTAATTAATTA 20686  
Qy 1789 ATTTTGAAC---TCCATGAATCATGAATGTTCTTAATGGAACCTGGAATGGAATCT 1845  
Db 20685 ATTTTCACTCTTCTCATGATCATGAATTTTAATTAATGACATCTGAATGGAATCT 20626  
Qy 1846 TTCCAAAGGTTTCAATTTACTTAATGATCAGATTCATTCATCAGAGATTCACCTTTCA 1905  
Db 20625 TTCCAAAGGTTTCAATTTACTTTGCTCCAGACTCAT-----CAGGAATCATCATCTA 20571

Qy 1906 TGCAGTATATAGCTTATGAAATGATTTCTTCAATATATAGCTGAAAGTTGAATTA 1965  
Db 20570 GGGCAGCTACAGCTTACCAAAATATATTTCTTATGATGAAAGACTGAAAGTCAAAATTA 20511  
Qy 1966 CTCTTGTATTCATTTTCTGCAAAATATGATGTTGTG 2000  
Db 20510 CTCTTGTATTCATGAGCTGTGAAATGATGTTGGG 20476  
RESULT 10  
AD97331\_3  
Continuation (4 of 6) of AD97331 from base 30001 (Human cancer associated sequence HD  
WP Sequence split into 6 fragments LOCUS AD97331 Accession Ad97331  
WP Fragment Name Begin End  
WP AD97331\_0 10000 11000  
WP AD97331\_1 10001 21000  
WP AD97331\_2 20001 31000  
WP AD97331\_3 30001 41000  
WP AD97331\_4 40001 51000  
WP AD97331\_5 50001 523643  
Query Match 24.4%; Score 488; DB 12; Length 110000;  
Best Local Similarity 75.0%; Pred. No. 1.1e-73;  
Matches 789; Conservative 0; Mismatches 205; Indels 58; Gaps 12;  
Qy 966 GGGCTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1025  
Db 92170 GGTAAAT 92229  
Qy 1026 TATATCAATATATGCAAGAGTATATCAATTAAGTATGATGATGATGATGATGATGAT 1085  
Db 92230 TATGCA-----ATATCAATATATATATATATATATATATATATATATATATAT 92274  
Qy 1086 TCCAGTGCATATTAAGTTTGTCTTATATCACTGTATGCTGTATTAAGTGTCAATAT 1145  
Db 92275 TCGAGTGCATATTAAGTTTGTCTTATATCACTGTATGCTGTATTAAGTGTCAATAT 92334  
Qy 1146 GTTATGTCTAAATAAACATATCTTATTTTAAATGCTTTATTAATAATAATATCTAA 1205  
Db 92335 ATTTATGTCTAAATAAGTA-----CTTAATTTTAAATAATTTTATGCAAAAAGATCTAA 92389  
Qy 1206 CAATCATTTGACATTCAGT 1265  
Db 92390 AATATATCTGAGCTTTCGAAAGTATATATTTTGTGTGTGTGTGTGTGTGTGTGTGT 92448  
Qy 1266 GATGATGAT-----CGGGGTGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1314  
Db 92449 GTTATGAGCTACGACTAATCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 92508  
Qy 1315 TTCTTAAACAA--CAGTGAATTTGCAATATCATGTTGACTTCTCTTATGAAAGATT 1372  
Db 92509 TTCTTAAATAATGAGCAATGAAATTTGCCACATTAATGCT-----CTTCAAAAGGATT 92564  
Qy 1373 TCTCTCTAGT 1432  
Db 92565 TCTCTGTACATATCAATGCGGTTTATGATGATTTTATCCACAGTGAACCTTCTTCAAAA 92624  
Qy 1433 TTGGA--TCAATCTCTCAAAACCTGTCTGTCTTAAACAACCTTAATTAATATTTCT 1491  
Db 92625 TTGGAATTAATCTCTCAAAACCTGTCTGTCTTAAACAACCTTAATTAATTTCT 92684  
Qy 1492 GAATCATTTGT 1551  
Db 92685 AATCTTTGT 92744  
Qy 1552 TCATTTCTGAGATGAAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1611  
Db 92745 TCA---AGAACTATCTATCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 92800  
Qy 1612 TTATCATGAGATTGCAAGATATACATGATGCTTCAAGGCTCACTTCACTTTAATTC 1671  
Db 92801 TTATCTGCGATGTCAGCAACTGTGTCAATCTTCAGG-----CTTATCTCAATTC 92855



PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251900P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-541565/60.  
XX  
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
XX  
PS Disclosure; SEQ ID NO 12875; 1701bp + Sequence Listing; English.  
XX  
XX  
XX The invention relates to novel genes (ABA11004-ABA21533) and proteins  
XX (ABA14678-ABAB18001) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
XX cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
XX infectious diseases such as viral, bacterial, fungal and parasitic  
XX infections. Note: The sequence data for this patent did not form part of  
XX the printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 21224 BP; 6415 A; 4173 C; 4443 G; 6193 T; 0 U; 0 Other;  
Query Match 24.0%; Score 480.2; DB 5; Length 21224;  
Best Local Similarity 73.9%; Pred. No. 2.3e-72;  
Matches 713; Conservative 0; Mismatches 223; Indels 29; Gaps 7;  
QY 1043 AAGTGATATCACAAATAAAGTACACAAAGCTTTTGGCTCCAGTGCATATATATAA 1102  
DB 14728 AAGCAAGTATCACAAATAAAGTACACAAAGCTTTTGGCTCCAGTGCATATATAA 14669  
QY 1103 GTTTGCTTAACTACACTGATGCTGTTAAAGTGCATATATGTTATGCTTAAATAAC 1162  
DB 14668 GTTAGGCTTACACTATGATGCTTAAATGTAATATGATATATGCTTTTAAATAT 14609  
QY 1163 A-----CATACCTTAATTTTAAAGCTTATATCTTAAATAATGCAATCATTTGG 1217  
DB 14608 AATGACATGATGCTTAAATTTTAAATATCTTATATGCTTAAATAATGCTTAAATATG 14549  
QY 1218 CATTCAGTATGTTATCTTTTGGCTGGAAGGCTTTTCTTATGATGACTGATCG 1277  
DB 14548 TCTTCACTGATGCTTAAAGGCT-----TGCTCAATGTTAGCGCTCTGATCATCA 14494  
QY 1278 GGAGTCAAGTCTGAAGCTTAAAGGCTGAGGCTGAGGCTTTTCTTAAACAAAGTAAAT 1337  
DB 14493 GAGTATGATGCTGAAGGCTTAAAGGCTGAGGCTGAGGCTTTTCTTAAACAAAGTAAAT 14434  
QY 1338 GCAATATCAGTATGCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1397  
DB 14433 TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 14374  
QY 1398 ATAGCATTTTATGACAGATGAGAACTTCTTGAATAATGATGATGATGATGATGATGAT 1457  
DB 14373 AAGCATTTTATCTACAGAGAACTTCTTGAATAATGATGATGATGATGATGATGATG 14314  
QY 1458 TCTGCTTAAACAACTAATGATTAATATATATCTGATGATGATGATGATGATGATGAT 1517  
DB 14313 TCTGCTTAAACAACTAATGATTAATATATATCTGATGATGATGATGATGATGATG 14254  
QY 1518 TTTTCAAGTCTTCAACGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1577  
DB 14253 GTTCAAGATCTTGGCAGGATGATGATGATGATGATGATGATGATGATGATGATG 14198  
QY 1578 CATTCATTAAGAAATCTCATCTGTTCAAGTTTATCATGATGATGATGATGATGATG 1637  
DB 14197 GATCTATCAAGAACTCTTATCATTTTATCATTTATGATGATGATGATGATGATG 14138  
QY 1638 TCATGCTTCAAGCTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1697  
DB 14137 TCATGCTTCAAGCTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 14083  
QY 1698 TGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1757  
DB 14082 TGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 14023  
QY 1758 CGACTTCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAA 1814  
DB 14022 CAATCTTCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCAAAATCTTCA 13963  
QY 1815 GTTCTTAAATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1874  
DB 13963 GTTCTTAAATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13903  
QY 1875 AGATCATTCATCCAGAGATCATCTTCAATGCAATGATGATGATGATGATGATGATG 1932  
DB 13902 AGATCATTCATCCAGAGATCATCTTCAATGCAATGATGATGATGATGATGATGATG 13848  
QY 1933 TTTCTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1992  
DB 13847 TTTCTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13768  
QY 1993 ATGTT 1997  
DB 13787 ATGCT 13783

RESULT 12  
AAL04361/c  
ID AAL04361 standard; DNA; 32192 BP.  
XX AAL04361;  
AC  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 7049.  
XX  
KW Human; reproductive system related antigen; reproductive system disorder;  
cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
PN WO200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205151P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215335P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234984P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246529P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.

PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251899P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI; 2001-465570/50.  
 XX  
 PT Isolated nucleic acid molecule encoding a reproductive system antigen is  
 PT used in preventing, treating or ameliorating a medical condition.  
 XX  
 PS Disclosure; SEQ ID NO 7049; 1297bp + Sequence Listing; English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC including cancer. The present sequence is a genomic sequence encoding a  
 CC protein of the invention  
 XX  
 SQ Sequence 32192 BP; 9426 A; 6338 C; 6511 G; 9917 T; 0 U; 0 Other;  
 Query Match 24.0%; Score 480; DB 4; Length 32192;  
 Best Local Similarity 73.9%; Pred. No. 2.5e-72;  
 Matches 776; Conservative 0; Mismatches 220; Indels 54; Gaps 11;

QY 971 AGATGTGACAGGATCTTGGAGATACGTCGCTTGGTTCACACCAAAATA 1030  
 DB 8208 ACAAATATACAGGACCTCAAGATATAGGCTTCAATTCAGTAATGAAATAAG 8149  
 QY 1031 CAAATATGACAGAGTATATCACAATTAAGTACACAGAGCTTTGGCTTCCA 1090  
 DB 8148 TGAATAT-----TGTAAATTAAGCAATTTGTGTACGTAATGCTTGGTTCCA 8099  
 QY 1091 GTGCATATTAAGTTTGTCTTACTACTAGTCTGTCTTAAGTGTCAATAGTCTT 1150  
 DB 8098 GTGCATATTAAGTTTATATCTACACTAGTCTGTCTTAAGTGTCAATAGTCTT 8039  
 QY 1151 GTCATAAAACACATACCTTAATTTAAATGCTTATTAACCTAAATAATGCTAACATC 1210  
 DB 8038 ATCTAAATAA-----TGTACATTAATTTAAACATTAATTAATTAACCTAAATAATGCTAACATC 7984  
 QY 1211 ATTTGAGCATTCAGTATGTAATCTTTTGTCTGGTGAAGAGCTTTTCTTATGATGA 1270  
 DB 7983 ATCTGAGCTTTAGTATGTAATCTTTTGTCTGGTGAAGAGAGCTTTCTTAAAGTTGA 7924  
 QY 1271 CTGAT-----CGGGGGTCAAGTGTGAAGCTTAAGGGTGTGGAGAGTTTCTT 1319  
 DB 7923 TGGGCTGCTGACTGATCAGGAGGTGTGGCTGTGAAGGTGTGGCTGTGGCAATTTTAA 7864  
 QY 1320 AAA-----ACAACAGTGAAGATTCGAATTCAGTTCCTTCTTCTTCAAGAAATTC 1374  
 DB 7863 AAAAAGAGACACACAGAAAGTTTGCACATCAATTAATGATCTTTTAAAGAAATTTTC 7804  
 QY 1375 TCTCTAGTGTGATGCTTTTGTATGACATTTTATGACAGTAAGAACTTCTTGAAT 1434  
 DB 7803 TCTGAAGCTGTGAGTCTTTTGTATGATTTTAACTTGAAGTGAATCTTCTTCAAAAT 7744  
 QY 1435 GGA-TCAATCTCTCAAAACCTGCTGTCTTAAACACTTAAGTTATATATTTCTGA 1493  
 DB 7743 GGAAGTCATCTCTCCAAACCTGCTGTCTTATCACTTAAGTTATATATTTCTGA 7684

QY 1494 ATCCATTTGTTCATTTCAACATTTTTCACAGTCTTTCACAGAGTATGATTCATCTC 1553  
 DB 7683 ATCCCTTCTGTTCATTTTCAACATTTTTCACAGTCTTTCACAGAGTATGATTCATCTC 7624  
 QY 1554 ATTTCCGATGATGAAATCTTTGCTCATTCATTAAGAAATTTCTCATCTGTTCAAGTT 1613  
 DB 7623 AAGAAATCACGT-----TCTTTCACATTCATTAAGAAATTTCTCATCTGTTCAAGTT 7568  
 QY 1614 TTTCTGATGATGACAGCAATTAAGTATGCTTTTCAAGGCTCTACCTTCACTTTAATTCGAG 1673  
 DB 7567 TATCATGAGTGGCAGCAATTTTCACTCATCTTTCAGG-----TTCCACTTATGATTTCTGA 7513  
 QY 1674 TTTCTGCTGTTTCTTACACATCTGTGTTCTTCTTCTCATTTGAAGTCTTGAACCTCTC 1733  
 DB 7512 TTTCTGCTGATTTTTCACATCTGTGCTGCTGCTTCTTCTCA-----CTGAATCCCTC 7461  
 QY 1734 CAAGTCATCCATGAGGTTGGAATGACCTTCTTCAAAATTTCTGTTAATTAATTTT 1793  
 DB 7460 AAAGTCACCTATGAGGATTTGGAATCACTTCTTCAAACTCTGTTAATGTTGATGATAT 7401  
 QY 1794 GA---CTTCCCATGATATGATATGTTCTTATATGACCTGGAATGCTTTCCTCA 1850  
 DB 7400 GACATCTTTCATGATATGATATGATATGTTCTTATGATGATGATGATGATGATGAT 7341  
 QY 1851 AAAGTTTTCATTTTACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1910  
 DB 7340 GAAGTTTTCATTTTACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7283  
 QY 1911 GTTATAGCTTATGATATGATATTTCTTCAATTAATGAGCTTGAAGTTGAATTTCTCT 1970  
 DB 7282 GGTATAGCTTATGATATGATATTTCTTCAATTAATGAGCTTGAAGTTGAATTTCTCT 7223  
 QY 1971 TGATCATTTTTCGAAATTAATGATTTGTG 2000  
 DB 7222 TGATCATGAGGCTGAGATGATGATGATGATG 7193

RESULT 13  
 ACN44010/c  
 ID ACN44010 standard; DNA; 198522 BP.  
 XX  
 AC ACN44010;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human genomic sequence hCG1643869.  
 XX  
 KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003073826-A2.  
 XX  
 PD 12-SEP-2003.  
 XX  
 PF 28-FEB-2003; 2003WO-US006235.  
 XX  
 PR 01-MAR-2002; 2002US-00087192.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW;  
 XX  
 DR WPI; 2003-328604/31.  
 XX  
 PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 PT comprises a nucleotide sequence.  
 XX  
 PS Claim 1; SEQ ID NO 244; Opp: English.  
 XX  
 CC The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable

of binding to Carcinoma Associated Protein (CAP); (iii) for screening of a bioactive agent capable of modulating the activity of CAP; (iv) for evaluating the effect of a candidate carcinoma drug; (v) for diagnosing carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biobchip; (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for determining Carcinoma Associated (CA) gene copy number. In addition, the CA genes are useful as DNA vaccines and the CAP are useful as markers of carcinoma including lymphoma. The present sequence is one such CA coding sequence. Note: This patent is an equivalent to basic patent US2002182586A1, for which no sequence data was published

Sequence 198522 BP; 58663 A; 35063 C; 36453 G; 68323 T; 0 U; 20 Other;

Query Match 24.0%; Score 479.6; DB 11; Length 198522;

Best Local Similarity 70.9%; Pred. No. 3.1e-72; Matches 812; Conservative 0; Mismatches 279; Indels 55; Gaps 11;

```

QY 877 TTTATGCTATGCTTAATATACATTAAGTTTGTGTTGTTTGTCTTATCTTATCTGTT 936
DB 131071 TATAAGTGAATAATATAATATCTGATCTACTAAATTTTCTCAGCTCAGATTC 131012
QY 937 GCGTGGGCTGAGCAAACTTTCTGTAAGGCTAGATGTCAGGCAATCCTTGAGAT 996
DB 131011 ACCAAAGATTAAAAATGCTAAGATACAGGAAAAATATACAGGCTCCGAGAGAC 130952
QY 997 ACTGAGGTTGGTTCATACCAACCAATATATACAGAAAGTGAATATACCA 1056
DB 130951 ATTGCAAGTTGGTTCAGACCAACCCCAATAG-----GAGAAATCATTA 130907
QY 1057 ATAAAGTAGTCACACAGCTCTTTGGCTCCAGTGCATATATAAGTTTGTCTTATCT 1116
DB 130906 ACAAGTAGTCACATGATTTTGGTTCCCAACATATATAAGTATGTTTACCT 130847
QY 1117 ACACGTATGCTGTTAAGTGTGCATAGTGTATGCTTAATAAAC-----ACATACCTT 1171
DB 130846 ATACTGAGCTATTAAGGAGGCAACGATTAATGCTCAAAAGCCAAATGTTATATACCTT 130787
QY 1172 AATTTAAAGTCTTATTAATCAAAAATGCTAACATCATTTGAGCATTCAGTAGTTG 1231
DB 130786 AACTTGAAATAATTTTATTTGCTTAATAAAGGTGAATGATCATCTGAGCCTTCAGTAGT-- 130729
QY 1232 TAATCTTTTGGCTGAGAGGCTCTTTCTTATGTA-----TGACTGATCGGGGCTCA 1284
DB 130728 CACTCTCTTGGCTGAGAGGCTCTTGCCTCATGATGATGCTGATCAGGGTGGT 130669
QY 1285 GGTGCTGAAGCTTAGAGGCTGCTGCTGCACTTTCTTAA-----ACAACGTGAAGATTGC 1339
DB 130668 GGTGCTGAAGCTTAGAGGCTGCTGCTGCACTTTCTTAAATAAGACACCATGAAAGTTTGC 130609
QY 1340 AATATCAGTGACTCTCTCTTCAATGAAGATTCTCTAGTGTGATGCTTTTGTAT 1399
DB 130608 TCTATCATGTGACTCTCTCTTCAATGAAGATTCTCTGCGATTAAGAGCTCTTTGAT 130549
QY 1400 AGCATTTTATGACAGTAGACCTCTTTGAAAAAT--GATTCATCTCTCAAAACCTGCT 1458
DB 130548 AGCATTTTATCCACATATAAATCTCTTCAAAATGGGGTCAAGCCTCTCAAACTCTGCC 130489
QY 1459 CTGCTTTAACAACCTTAAGTTAATATATATCTGAATTCATGTTGTCATTTCAACATTT 1518
DB 130488 GCTGCTTACGACCTAAGTTTATTTAATATCTTAACCTCTGTTGTCATTTCAACAAG 130429
QY 1519 TTCACAGTCTCTTACCCAGAGTAGTGCATCTCATTTTCTGAGATGAATCTTTGCTC 1578
DB 130428 TTCACAGCATCTTAACTTAAGAAATGTTCTATCTCAAGAAACCACTT---TCTTTGCTC 130373
QY 1579 ATCCATAAGAAATAATCTCTCATCTGTTCAAGTTTATCATGATGAGATGAGCAATACAT 1638
DB 130372 ATCCATAAATGCAATCTCTCATCTGTTCAAGTCTCTCATGATGAGATGAGCAATTAAT 130313
QY 1639 CATGCTTCAGGCTCATCTTCAATTTTCAAGTGTCTTCTGCTGTTTCTACACATCT 1698
DB 130312 CACGCTTAAAG-----CTCCACTTCTATATTAAGTTCTCTTATTAATTTCCACACAGCT 130260

```

```

QY 1699 GTGGTCTTCTCTCATTTGAAGTCTTGAACCTTCCAAATGTCATCC-ATGAGGTTGGAT 1757
DB 130259 ACAGTACTTCTCTCAGTGAACCTTAACTCCCTCAGAGTCATCAATGAGGATGAGAT 130200
QY 1758 CGACTTCTTCCAAATTTCTGTTATATTTAATTTGA---CTGCCATGAATCATGAT 1814
DB 130199 CAATGACTTCCAACTCCCTGTAAGTGTGAATTTCTGACTCTCCCATGATCATGAT 130140
QY 1815 GTTCTTAATGACCTGGAATGTGAATCTCTTCCAAAGGTTTTCATTTACTTACTGTC 1874
DB 130139 GTTCTTAATGACCTGTAAGACGTAATCTCTTCCAGAGGTTTAAATGCACTGTGTC 130080
QY 1875 AGATCCATTCATCCAGAGATCCACTTCAATGCAAGTTATAGCCTTAATGGAATGATTT 1934
DB 130079 AGATGCAAT-----CAGAGACTCCGCTATCTATAGCAGTTATAGCCTTGTAAGTACTT 130025
QY 1935 CTTCAATTAATAGCTTGAAGTTGAATTAATCTGATTCATTTCTGCAAAATGAT 1994
DB 130024 CTTAAGTATATGAGACTTGAAGTGAATTAATCTCTTATTCATTTGATGATGCAATGAT 129965
QY 1995 GTTGTG 2000
DB 129964 GTTTTG 129959

```

```

RESULT 14
ADA53284/C
ID ADA53284 standard; CDNA; 2448 BP.
XX
XX ADA53284;
XX
XX 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 852.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
XX Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
XX Homo sapiens.
XX
XX EPI293569-A2.
XX
XX 19-MAR-2003.
XX
XX 21-MAR-2002; 2002EP-00006586.
XX
XX 14-SEP-2001; 2001JP-00328381.
XX
XX 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) RES ASSOC BIOTECHNOLOGY.
XX
XX (RENS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Iehi S,
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI. 2003-395539/38.
XX
XX F-PSDB; ADA54923.
XX
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
XX and/or membrane proteins, useful for developing medicines for diseases in
XX which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 1; SEQ ID NO 852; 205BP; English.
XX
XX The present invention relates to novel human secretory or membrane
XX proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
XX ADA54071). The coding sequences are useful in the gene therapy of
XX diseases caused by abnormalities of the proteins, e.g. cancer,
XX inflammatory diseases, osteoporosis or neurological disease.
XX

```





```
Db 39818 CATTGCTGTTGATGATTTTAAACCAAGTATTAATATCTGATCATTTGTTG 1504
Oy 1445 TCTCAAAACCTGCTGCTTAAACCAAGTATTAATATCTGATCATTTGTTG 1504
Db 39758 TCTCAAAACCTGCTGCTTAAACCAAGTATTAATATCTGATCATTTGTTG 1504
Oy 1505 TCATTTCAACAATTTTCAAGTCTTCAAGAGTATTCATCTCATTTTCTTGA 1564
Db 39698 TCATTTCAACAATTTTCAAGTCTTCAAGAGTATTCATCTCATTTTCTTGA 1564
Oy 1565 TCGAATCTTTGCTCATTAAGAAGAAATTCATCTGTTCAAGTTTATCATGAAT 1624
Db 39638 T-----TCTTGGCTCAT-CATAAGAAACAACCCCTCATCTCAAGTTGATCATGAAGT 1624
Oy 1625 TCGACCATATCATGATGCTTCAAGGCTTCAAGTCTTCAAGTCTTCAAGTCTTCA 1684
Db 39583 TCGACCATATCATGATGCTTCAAGGCTTCAAGTCTTCAAGTCTTCAAGTCTTCA 1684
Oy 1685 TTTTACCAACATCTGCTTCTTCTTCTTCAATGAAGTCTTGAACCTTCAAGTCTCA 1744
Db 39528 TTTTCCCGGCAATCTGCTTCTTCTTCTTCAAGTCTTGAACCTTCAAGTCTCA 1744
Oy 1745 TGAAGGTTGAATGATCTTCTTCAAAATCTTCTTCAATTAATTTTGA---CTGCTC 1801
Db 39468 TGAAGGTTGAATGATCTTCTTCAAAATCTTCTTCAATTAATTTTGA---CTGCTC 1801
Oy 1802 ATGAATCATGATGTTCTTAAATGAAGCAAGTCTTCAAGTCTTCAAGTCTTCA 1861
Db 39408 ATGAATCATGATGTTCTTAAATGAAGCAAGTCTTCAAGTCTTCAAGTCTTCA 1861
Oy 1862 ATTTACTAGTCCAGATCCATCCATCCAGAGATCCATCTTCAAGTCTTCAAGTCTT 1921
Db 39348 TTTTCAATTAATTTGCTTCAAGTCTTCAAGAGATCCATCTTCAAGTCTTCAAGTCTT 1921
Oy 1922 ATGAATGATTTCTTCAATTAATGAAGTCTTGAAGTCTTCAAGTCTTCAAGTCTT 1981
Db 39288 ACAAATGATTTCTTAAATAAATAAGTCTTGAAGTCTTCAAGTCTTCAAGTCTT 1981
Oy 1982 CTGCAAAATGATGTTGTTG 2000
Db 39228 CTGTAAGTGAAGTGTGTTG 39210

RESULT 16
AAD46127
ID AAD46127 standard; DNA; 163350 BP.
XX
AC AAD46127;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human tumour suppressor gene.
XX
XX Human; tumour suppressor protein; cell proliferative disorder; vaccine;
XX KM inflammation; brain cancer; adenocarcinoma; cervix cancer; bone cancer;
XX KM apoptosis; leukaemia; lymphoma; melanoma; therapy; chromosome 13; gene;
XX ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 2001.160511
XX FT /tag= a
XX FT /product= "Human tumour suppressor protein"
XX FT exon
XX FT 2001.2178
XX FT /tag= b
XX FT intron
XX FT 2179.159807
XX FT /tag= c
XX FT exon
XX FT 159808.160511
XX FT /tag= d
XX FT
XX PN WO200268468-A2.
```

```
XX 06-SEP-2002.
PD 05-FEB-2002; 2002WO-US003235.
XX 27-FEB-2001; 2001US-00793706.
XX (PEKE ) PE CORP NY.
XX Gong F, Yan C;
XX WPI; 2002-698658/75.
XX P-PSDB; AAB28633.
XX New human tumor suppressor proteins, useful for developing human
XX therapeutic agents, or preventing or treating inflammation, or disorders
XX associated with cell proliferation, e.g. bone cancer, brain cancer,
XX leukemia or lymphoma.
XX Claim 4; Fig 3; 200pp; English.
XX The invention relates to human tumour suppressor polypeptides and
XX polynucleotides. Sequences of the invention are useful for diagnosing,
XX preventing or treating inflammation, or disorders associated with cell
XX proliferation and apoptosis e.g. bone cancer, brain cancer, cervix
XX cancer, adenocarcinoma, leukaemia, lymphoma or melanoma. They are
XX particularly useful as models for developing human therapeutic targets,
XX identifying therapeutic proteins, or serving as targets for the
XX development of human therapeutic agents that modulate tumour suppressor
XX protein activity in cells and tissues that express the tumour suppressor
XX protein. Polypeptides of the invention are used for identifying agents
XX that modulate their activity. They are useful for raising antibodies or
XX eliciting an immune response; as a reagent in assays designed to
XX quantitatively determine levels of the protein (or its binding partner or
XX ligand) in biological fluids; or as markers for tissues in which the
XX corresponding protein is preferentially expressed. The invention is also
XX used as vaccines. The present sequence is human tumour suppressor gene
XX located on chromosome 13
XX
XX Sequence 163350 BP; 40430 A; 36356 C; 37002 G; 39020 T; 0 U; 10542 Other;
XX
XX Query Match 23.5%; Score 469.4; DB 6; Length 163350;
XX Best Local Similarity 75.0%; Pred. No. 1.7e-70;
XX Matches 779; Conservative 0; Mismatches 201; Indels 59; Gaps 13;
XX
Oy 978 ACAGGATATCTTGGAGATCTGAGTGTGTTGTTCCATACACCAATTAATACAAATAT 1037
Db 100173 ACAGGATATCTTGGAGATCTGAGTGTGTTGTTCCATACACCAATTAAT 1037
Oy 1038 GCAAGAGTGAATATCAATTAAGTGAAGTCAAGTCTTGGTCCAGTGCAAT 1097
Db 100223 ----AAGCAATATCAATTAAGTGAAGTCAAGTCTTGGTCCAGTGCAAT 1097
Oy 1098 TAAAGTTTGGTTA---TACTACCTGTAAGTCTTGAAGTGAATTAAGTGAAT 1154
Db 100278 TAAAGTTTGGTTA---TACTACCTGTAAGTCTTGAAGTGAATTAAGTGAAT 1154
Oy 1155 AAAAAAAC-----ACATACCTTAATTTTAAATGCTTATTACTTAAATAATCTAACAT 1209
Db 100338 GAAAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1209
Oy 1210 CATTGAGCATTCAGTGAAGTGTGTAATCTTTTGTGTTGAGTCTT-----TTC 1260
Db 100397 CATATGAGCTTCACTGA---GTATCTTTTGTGTTGAGTCTT-----TTC 1260
Oy 1261 TTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
Db 100454 TGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
Oy 1321 AA-----ACAAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1375
Db 100514 AATTAAGCAATTAAGTGTGTTGCTGATGATGATGATGATGATGATGATGATGAT 1375
```

```
QY 1376 CTCAGTGTGTGATGCTTTTGTATGACATTTATGCAAGATGACCTCTTGGAAATATG 1435
DB 100574 CTGTACATTCACAG---TTGTATGACATTTTACCCAGCGAGAAATGCTTTCAAAATGG 100630
QY 1436 GA-TCAATCTCTGCAAAACCTGCTGCTTTAAACAACCTAATATATATATTTCTGAA 1494
DB 100631 GAGCCAGTCTCTCAAAACCTGCGACGCTTTATCAATAATTTATGCAATAGTCTAAA 100690
QY 1495 TCCATTTGTTGATCTTCAACAATTTTACAGTGTCTTCCACGAGATAGATTCATCTCA 1554
DB 100691 TCTCTGTTGTGATCTTCAACAATTTTACAGATCTTCCACGAGATAGATTCATCTCA 100750
QY 1555 TTTCTGAGATGGAATCTTGTGCTCATCTCAATAGAAATAATTTCTCTATCTGTTCAAGTTT 1614
DB 100751 AGAAGACACTT---TCTTGTGCTCATCTCAATAGAAATAATTTCTCTATCTGTTCAAGTTT 100806
QY 1615 ATCATGAGATTTGACGAATACAGTCAATGCTTTCAGGCTCACTTCACTTTATTTCCAGT 1674
DB 100807 ATCATGAGATTTGACGAATACAGTCAATGCTTTCAGGCTCACTTCACTTTATTTCCAGT 100864
QY 1675 TCTCTGCTGTTTCTTACACATCTGTGTCTCTCTCTCATGTAAGTCTTGAACCTCTCC 1734
DB 100865 TCTCTGCTGTTTCTTACACATCTGTGTCTCTCTCTCATGTAAGTCTTGAACCTCTCA 100924
QY 1735 AAGTCATCCATGAGAGGTTGGAATCGACTTCTTCCAAATTCGTATATATTTATTTTG 1794
DB 100925 AAGTCATCCATGAGAGGTTGGAATCGACTTCTTCCAAATTCGTATATTTATTTTG 100984
QY 1795 ACC---TCCCATGATCATGAAATGTTCTTAAATGCACTGGAATGATGCTTTCGA 1851
DB 100985 ACCATATCCCATGATCATGAAATGTTCTTAAATGCACTGGAATGATGCTTTCGA 101044
QY 1851 AAGGTTTAAATTTACTAGTCAATCCATCCATCCATCCAGAGATCCACTTTCAATGCCG 1911
DB 101045 AAGGTTTAAATTTACTAGTCAATCCATCCATCCATCCAGAGATCCACTTTCAATGCCG 101099
QY 1912 TTAATGACCTTATGGAATGATTTCTTCAATATTAAGGCTTGAAGTGAATTTACTGCT 1971
DB 101100 TTAATGACCTTATGGAATGATTTCTTCAATATTAAGGCTTGAAGTGAATTTACTGCT 101159
QY 1972 GATCCATTTTCTGCAAAAT 1990
DB 101160 GATCCATTTTCTGCAAAAT 101178
```

## RESULT 17

AAL52246.1

Continuation (2 of 4) of AAL52246 from base 100001 (Human genomic DNA for the gene encod

WP Sequence split into 4 fragments LOCUS AAL52246 Accession AAL52246

WP	Fragment Name	Begin	End
WP	AAL52246_0	1	110000
WP	AAL52246_1	100001	210000
WP	AAL52246_2	200001	310000
WP	AAL52246_3	300001	378361

Query Match 23.3%; Score 466.6; DB 10; Length 110000;

Best Local Similarity 72.7%; Pred. No. 4.9e-70;

Matches 758; Conservative 0; Mismatches 234; Indels 51; Gaps 10;

```
QY 977 TACAGGCAATCTTGGAGATCTGCGGTTTGGTTCCATACACCAACATATACAAATA 1036
DB 85373 TACAGGCAATCTTGGAGATCTGCGGTTTGGTTCCATACACCAACATATACAAATA 85432
QY 1037 TGCAGAGATGATATCACAATTAAGTAGTCAACAAGTCTTTGGCTTCCAGTGCAT 1096
DB 85433 TGG-----CAATTAAGCAAGTCAATATAGACTGTTTGTCTCCAGTGCAT 85477
QY 1097 AATAAGTTTGGCTTATATACATCTAGTCTGTTAAGTGTGCAATAGTGTATGCTAA 1156
DB 85478 AATAAGTTTATTTTACACTCTAGTGTGCT---TAAAGTGTCAAGTCAATATGCTTT 85534
QY 1157 AAAAAACATACCT-TAATTTTAAATGCTTATTTACTAAAAATGCTAACATCATTTG 1215
```

```
DB 85535 AAAAAATGATATCTTAATTTTAAATATATTTTACTGCAAAAAATGCTAACACGCTCCG 85594
QY 1216 AGCATTCAGTGAAG---TTGTATCTTTTGTGCGGAGAGGCTTTCTTATGATGACT 1272
DB 85595 AGCTTCAGAGAGTCAATAATTAATTTTGTGATGAGAGGCTTGCCTCAATGTTGAAG 85654
QY 1273 GAT-----CGGGGGTCAAGGCTGGAAGCTTAAAGGAGCTGTGCAAGTTCTTAA 1321
DB 85655 GCTGCTGACTGATCAAGAGAGATATGTTGCTGAAGGTTTGAATGAGCTGTGAAATTTCTTAA 85714
QY 1322 AACCAACAGTGAAGATTTGCAATATCAAGTGAATCTTCTCTTCAATGAAGATTTCTCTAG 1381
DB 85715 AATAAGATATACATGAATATTTGTCAACTGACTTCTCTTCAATGAAGATTTCTCTAG 85774
QY 1382 TGTGTGATGCTTTTGTATGCAATTTTATGCAAGTGAAGCTTTTGAATAATG-GATCA 1440
DB 85775 CATGCAATGCTATTTGATGATGCAATCTTAATCAAGTGAAGCTTTTCAAAATTTGATCA 85834
QY 1441 ATCTCTGCAAAACCTGCTGCTTAAACAACCTAAGTAAATATATATCTGAATCATTT 1500
DB 85835 GTTCTCTTAAACCTGCTGCTGCTTATCACTAAATTTTGTATATTTCTAAATTTCTTG 85894
QY 1501 GTTGTCAATTTCAACAATTTTCAAGTGTCTTCAAGAGATGATTCATCTCATTTCTCT 1560
DB 85895 GTTGTCAATTTCAACAATGCTCATAGATCTTTACAGAGATGATTTTCATCTCA----AG 85950
QY 1561 GAGATGAATCTTTGTCTATCATTAAGAAATTCCTCATCTGTTCAAGTTTATCATG 1620
DB 85951 AATCACTATTTCTTGTCTTATATTAAGAAACATTTCTCATCTGCTCAAAATTTTAAATG 86010
QY 1621 AATATGAGCAATTAAGATATGCTTCAAGGCTCACTTCACTTAAATTTCCAGTCTCTT 1680
DB 86011 AATATGAGCAATTAAGATATGCTTCAAGGCTCACTTCACTTAAATTTCCAGTCTCTT 86065
QY 1681 GCTGTTTCTACACATCTGTGTCTTCTCTCATTTGAAGTCTGAACCTCTCAAGTCA 1740
DB 86066 ACTATTTTCCACATCTGTGACTACTCTCTCTCATTTGAAGTCTGAACCTCTCAAGTCA 86125
QY 1741 TCCATGAGGTTGGAATGCACTTCTTCAAAATTCCTGTTAATATTTATTTGACC--- 1797
DB 86126 TCCATGAGGTTGGAATGCACTTCTTCAAAATTCCTGTTAATATTTATTTGACC--- 86185
QY 1798 TCCCATGATCATGAAATGTTCTTAAATGCACTGGAATGGAATCTTCCAAAAGTT 1857
DB 86186 TCCCATGATCATGAAATGTTCTTAAATGCACTGGAATGGAATCTTCCAAAAGTT 86245
QY 1858 TTCAATTTACTAGTCCAGATCCATCCATCCAGAGATCCACTTCAATGCGAGTTATAG 1917
DB 86246 TTCAATTTACTAGTCCAGATCCATCCATCCAGAGATCCACTTCAATGCGAGTTATAG 86300
QY 1918 CCTTATGGAATGATTTTCTCAATATTAAGGCTTGAAGTGAATTAATCTCTTATATCA 1977
DB 86301 CCTTATGGAATGATTTTCTCAATATTAAGGCTTGAAGTGAATTAATCTCTTATATCA 86360
QY 1978 TTTTCTGCAAAATTAATGATGTG 2000
DB 86361 TGAAGTGAATGATGATGATG 86383
```

## RESULT 18

AD059207

ID AD059207 standard; DNA; 262090 BP.

AD059207;

09-SEP-2004 (first entry)

MSI-H carcinoma genomic DNA sequence SEQ ID NO:44.

coding mononucleotide repeat; CMNR; gene; antibody; MSI-H tumour;  
MSI-H carcinoma; high microsatellite instability tumour;  
high microsatellite instability carcinoma; cytostatic; ds.

OS Homo sapiens.  
 XX KR2004008012-A.  
 XX 28-JAN-2004.  
 XX 15-JUL-2002; 2002KR-00041304.  
 XX 15-JUL-2002; 2002KR-00041304.  
 XX 15-JUL-2002; 2002KR-00041304.  
 XX (KIMH/) KIM H G.  
 XX (KIMN/) KIM N G.  
 XX (LEEJ/) LEE J S.  
 XX (RHEE/) RHEE H S.  
 XX Kim HG, Kim NG, Lee JS, Rhee HS.  
 XX WPI; 2004-386326/36.  
 XX Genes containing coding mononucleotide repeats are useful in developing  
 PT an antibody against MSI-H (hugh (sic high) microsatellite instability)  
 PT tumor.  
 XX Claim 3; SEQ ID NO 44; 578bp; Korean.  
 XX The present invention describes genes containing coding mononucleotide  
 CC repeats (CNMRs). The genes are useful for the development of an antibody  
 CC against MSI-H (hugh microsatellite instability) tumour. Also described:  
 CC (1) cDNA genes containing CNMRs with 10 or more nucleotide sequences, and  
 CC selected from the cDNA genes having the nucleotide sequences of SEQ ID  
 CC NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35,  
 CC 37, 39, 41 and 43; (2) cDNA genes, which are frameshift mutated by  
 CC deletion or insertion of one or more base in the CNMRs; (3) genomic DNA  
 CC genes containing CNMRs with 10 or more nucleotide sequences, and selected  
 CC from the genomic DNA genes having the nucleotide sequences of SEQ ID  
 CC NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36,  
 CC 38, 40, 42 and 44; and (4) genomic DNA genes, which are frameshift  
 CC mutated by deletion or insertion of one or more base in the CNMRs. The  
 CC genes have cytosinatic activity. The present sequence represents an MSI-H  
 CC carcinoma genomic DNA sequence from the present invention.  
 XX Sequence 262090 BP; 74456 A; 48367 C; 51411 G; 86847 T; 0 U; 1009 Other;  
 SQ  
 Query Match 23.3%; Score 465.8; DB 12; Length 262090;  
 Best Local Similarity 73.1%; Pred. No. 6.9e-70;  
 Matches 779; Conservative 0; Mismatches 232; Indels 54; Gaps 12;  
 QY 972 GATAGTACAGCATACCTTGAGATACCTGTGGTTGGTTCCATACACCACCAATATAC 1031  
 DB 118291 GCTAATACAGGATACCTTGAGATACCTGTGGTTGGTTCCAGGCTACCCCAATAAAC 118350  
 QY 1032 AATATGCAAGAGTG-----AATACATTAAGTAGTACACAAAGTCTTT 1081  
 DB 118351 AATATGCAATTAACAGTAAATCAATTAATTAAGTAGTACACAAAGTCTTTT 118410  
 QY 1082 GGCTCCAGTGCATTAAGTTTGTATCTACTACCTGAGTGTCTGTTAAGTGCAA 1141  
 DB 118411 GGTTCCAGTGCATTAAGTTTGTATCTACTACCTGAGTGTCTGTTAAGTGCAA 118470  
 QY 1142 TAGTGTATGCTTAAGTAA-----CACATCTTAATTTTAAATGCTTAATTAAGTAA 1198  
 DB 118471 TAGCATTAATGCTTAAGTAA-----CACATCTTAATTTTAAATGCTTAATTAAGTAA 118530  
 QY 1199 ATGCTAACATCAATTTGAGCATCTGAGTGTATCTTTTGTGCTGAGAGTCTTT 1258  
 DB 118531 ATGCTAACATCAATCTGAACCTTCAGCAGCCAT-ATCTTTTGTGCTGAGAGTCTTTG 118589  
 QY 1259 TCTTATGATGATGATCGGGGGTCA-----GGTCTGAAGCTTAGGGTGCTG 1307  
 DB 118590 CTTCCATGTCAGTGGCTGATGATCAGATGGTGGTTACTGCAAGTGTGATTTG 118649  
 QY 1308 TGGCAGTTTCTTAAACAAACA-----GTGAAGATTGCAATATCAGTTGACTCTTCTTCA 1363

DB 118650 TGACAAATTTCTTAATAAATACAAACATGAGTTTGACATATCAGTTGACTCTTC--TTTC 118707  
 QY 1364 TGAAGATTTCTCTGATGTGTGATGCTTTTGTATAGCATTTTATGACAGTAGAACTT 1423  
 DB 118708 GTGAATGATTTCTCTGATGATGATGCTTTTGTATAGCATTTTATGACAGTAGAACTT 118767  
 QY 1424 CTTTGAATAATGGA-----TCAATCTCTCAAAACCTGCTGTGCTTTAACAACCTTAAGT 1478  
 DB 118768 CTTTCAAAATTTGAGATCGCTCTTCCCTCAAAACCTGCTGCTGCTTTATCAACTTAAGT 118827  
 QY 1479 AATATTAATTTCTGAATCCATGTTGTGATTTCAACAATTTTCAAGTCTTCAACGAG 1538  
 DB 118828 --TATGATGCGTAATTTTGTGTATTTCAACAGTGTTTATAGCTTCTTCAACGAG 118885  
 QY 1539 AGTAGATTCATCTCATTTCTCTGAGATGGAATCTTGTCTCATCAATGAAGAAATTCCT 1598  
 DB 118886 AGTAGATTCATCTCATTTCTCTGAGATGGAATCTTGTCTCATCAATGAAGAAATTCCT 118941  
 QY 1599 CATCTGTCAAGTTTATCATGAGATTCAGCATATCAGTATGCTTTCAGGCTCACTT 1658  
 DB 118942 CATCCAGTCAAGTTTATCATGAGATTCAGCATATCAGTATGCTTTCAGGCTCACTT 118996  
 QY 1659 CACTTAAATCCAGTCTCTGCTGCTTCTACCAATCTGAGTCTTCTGCTCACTTGA 1718  
 DB 118997 CTTATATATTTAATTTCTCTGCTGCTTATTAACATCTGCTGCTCACTTGA 119056  
 QY 1719 AGCTTGAACCTCTTCAAGTATCATGAGGCTTGAATGCACTTCTTCAAAATTCCTGT 1778  
 DB 119057 AGCTTGAACCTCTTCAAGTATCATGAGGCTTGAATGCACTTCTTCAAAATTCCTGT 119116  
 QY 1779 TAAATTTAATTTTGA--CCTCCATGAATCATGAATGTTCTTAATGCAACCTGAGAT 1835  
 DB 119117 TAAATTTAATTTTGA--CCTCCATGAATCATGAATGTTCTTAATGCAACCTGAGAT 119176  
 QY 1836 GGTGAATCTTCCAAAGTTTTCATTTACTTATGTCAGATCCATCCATGAGAT 1895  
 DB 119177 GGTGAATCTTCCAAAGTTTTCATTTACTTATGTCAGATCCATCCATGAGAT 119232  
 QY 1896 CCACTTGAATGCAAGTATAGCTTATGAGATGATTTCTTCAATTAATGAGCTTGA 1955  
 DB 119233 CCACTTGAATGCAAGTATAGCTTATGAGATGATTTCTTCAATTAATGAGCTTGA 119292  
 QY 1956 GTTGAATTTACTCTTGTATCCATTTTCTGCAAAATAGATGTTGTG 2000  
 DB 119293 GTTGAATTTACTCTTGTATCCATTTTCTGCAAAATAGATGTTGTG 119337  
 RESULT 19  
 ID ACN37213  
 ID ACN37213 standard; DNA; 23271 BP.  
 XX ACN37213;  
 AC 18-NOV-2004 (first entry)  
 XX 18-NOV-2004 (first entry)  
 DT Human periodontal disease related gene KRTHA2 SEQ ID NO:123.  
 DE periodontal disease; polymorphism; ds; human; gene; SNP;  
 KW single nucleotide polymorphism.  
 XX Homo sapiens.  
 OS  
 XX Location/Qualifiers  
 FH misc\_feature  
 FT 1796  
 FT /tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "Variable nucleotide T,C"  
 FT 1867  
 FT /tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "Variable nucleotide G,A"  
 FT 1964  
 FT misc\_feature  
 FT /tag= c

PT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "Variable nucleotide G,A"
PT		4942
PT	misc_feature	/*tag= d
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide T,C"
PT		5020
PT	misc_feature	/*tag= e
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide A,G"
PT		5133
PT	misc_feature	/*tag= f
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide A,T"
PT		5841
PT	misc_feature	/*tag= g
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide C,T"
PT		6113
PT	misc_feature	/*tag= h
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide T,C"
PT		6132
PT	misc_feature	/*tag= i
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide A,G"
PT		6164
PT	misc_feature	/*tag= j
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide T,G"
PT		6347
PT	misc_feature	/*tag= k
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide T,C"
PT		6380
PT	misc_feature	/*tag= l
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide C,T"
PT		7316
PT	misc_feature	/*tag= m
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide T,G"
PT		7905
PT	misc_feature	/*tag= n
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide C,G"
PT		8025
PT	misc_feature	/*tag= o
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide C,A"
PT		8095
PT	misc_feature	/*tag= p
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide A,G"
PT		8252
PT	misc_feature	/*tag= q
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide C,G"
PT		8405
PT	misc_feature	/*tag= r
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide A,G"
PT		8828
PT	misc_feature	/*tag= s
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide C,A"
PT		8940
PT	misc_feature	/*tag= t
PT		/standard_name= "Single nucleotide polymorphism"
PT		/note= "Variable nucleotide T,C"
PT		8985
PT	misc_feature	/*tag= u
PT		/standard_name= "Single nucleotide polymorphism"

FT		/note= "Variable nucleotide T,C"
FT	misc_feature	/tag= v 9606
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide T,C"	
FT	misc_feature	9763
FT	/tag= w	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide T,C"	
FT	misc_feature	9664
FT	/tag= x	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide G,A"	
FT	misc_feature	9799
FT	/tag= y	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide G,A"	
FT	misc_feature	9980
FT	/tag= z	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide T,C"	
FT	misc_feature	10001
FT	/tag= aa	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide G,A"	
FT	misc_feature	10072
FT	/tag= ab	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide C,G"	
FT	misc_feature	10105
FT	/tag= ac	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide G,A"	
FT	misc_feature	10326
FT	/tag= ad	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide deletion,C"	
FT	misc_feature	10920
FT	/tag= ae	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide T,C"	
FT	misc_feature	11619
FT	/tag= af	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide C,G"	
FT	misc_feature	11796
FT	/tag= ag	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide A,C"	
FT	misc_feature	12217
FT	/tag= ah	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide C,G"	
FT	misc_feature	12218
FT	/tag= ai	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide C,G"	
FT	misc_feature	12672
FT	/tag= aj	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide T,C"	
FT	misc_feature	12954
FT	/tag= ak	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide T,G"	
FT	misc_feature	13271
FT	/tag= al	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide G,A"	
FT	misc_feature	14010
FT	/tag= am	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	/note= "Variable nucleotide C,G"	

```
FT misc_feature 14011
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide C,G"
FT misc_feature 14508
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide A,C"
FT misc_feature 14594
FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide T,C"
FT misc_feature 14771
FT /tag= aq
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide C,T"
FT misc_feature 14772
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide A,G"
FT misc_feature 15043
FT /tag= as
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide G,A"
FT misc_feature 15126
FT /tag= at
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide C,G"
FT misc_feature 15212
FT /tag= au
```

```
Query Match 23.3%; Score 465.6; DB 13; Length 23271;
Best Local Similarity 73.1%; Pred. No. 6.9e-70;
Matches 768; Conservative 0; Mismatches 224; Indels 58; Gaps 11;
```

```
QY 972 GATAGTACGCGATCTTGAGATACCTGGGTTGGTCCATACCAACAATATATC 1031
DB 17926 GAAGATACAGGATACCTCGAGATAGTCAGTTCAGTTCAGACACCAATTAAC 17985
QY 1032 AATATGCAAGAGTGATATACAAATTAAGAGTACACAAGCTTTGGCTCCAG 1091
DB 17986 AATGTA-----ACATTAAGTACACACAACTTTTGGTTTCTG 18030
QY 1092 TGCATATAAAGTTTGTATATCTACCTGTAGTCTGTAAAGTGCATAGTGTATG 1151
DB 18031 TGTATGTAAACATTAATTTACCTATCTATAGTCTTATTAAGTACATAGCATTAAG 18090
QY 1152 TCTAAAAA---ACACATACCTTAATTTAAATGCTTTATTAATAAAATGCTAAC 1207
DB 18091 GCTTAAAAAATGTATATTAATTTAAAAATGCTTTAATGCTAAAAATGCTAAGG 18150
QY 1208 ATCATTTGAGCATCTGAGTGTATCTTTTGTGCGGAGAGGCTTTCTTTTGA 1267
DB 18151 ATCATCTGAGCTTCGCAAGTTATATCTTTTGTGCGGAGAGGCTTCCTCAATG 18210
QY 1268 TGAATGAT-----CGGGGTCAAGTGTGAGCTTGAAGTGGCTGTGGCAGTT 1316
DB 18211 TGAATGCTAGAGCATGATCAGGTTGATGATGCTGCAAGTATGCTGTGGCAGATT 18270
QY 1317 CTTAAA-----ACACAGTGAAGATTGCAATATCACTTGACTCTTCATGAAGAT 1371
DB 18271 CTGAAAATGAGACATCAGTGAAGTTCGCGCATCACTGCTCTTCATGAAGAT 18330
QY 1372 TTCTCTAGTGTGATAGTCTTTTATAGCATTTTATGACAGTGAAGTCTTTGAAA 1431
DB 18331 TTCCCTGTGACATGTATGCTGTATGACATTTTA---ACGTGAAATTTCTTCAAA 18387
QY 1432 A-TTGGATCAATCCTCTCAAAACCTGCTGCTTTTAAACAACCTAAGTAAATATATTC 1490
DB 18388 ACTAGGATAGTCTCTCAAAACCTGCTGCTTTTAACTAATGATGATGATGATTC 18447
QY 1491 TGAATCAATGTTTCAATTTCAAAATTTTCAAGTGTCTTCCACGAGATAGATTCAT 1550
```

```
DB 18448 CAATCCTTTGTGCCATTTGATGATGTTCAACAGATCTTCACGAGATAGTTCCAT 18507
QY 1551 CTCATTTCCGAGATGGAATCTTTGCTCATCATTAAGAAGAAATCTCATCTGTTCAAG 1610
DB 18508 CTCAGAAACAGACT---TCTTTGTATGAATTAAGAAGAGTTCCTCATCCGTTCC--G 18561
QY 1611 TTTTATCATGAGATTGACAGCAATATAGTCAATGCTTCAGGCTCACTTCACTTTAATTC 1670
DB 18562 TTTTGCATTAACATGGCAACATTCAGTCAATCTTCAGG----CTCCAAATCTTAATTT 18616
QY 1671 CAGTCTCTGCTGCTTTTACACATCTGTGGTCCCTCCATTAAGTCTTGAACCT 1730
DB 18617 TAAITCTCTGTGCTTTCCAAACATCTTAATGTTAATCTTCTGCTGAAGTCTTAAGCC 18676
QY 1731 CTCAGATCATCATGAGGTTGGAATGACTTCTCCAAATTCCTGTAAATATATAT 1790
DB 18677 CTCACATCATATCAGATGAGTTGGAATGACTTCTCCAACTCCTGTAAATGATAT 18736
QY 1791 TTTGACCTCCATGATATGATGTTCTTAATGCGACCTGGAATGATCCTTTCCA 1850
DB 18737 TTTGACCTCTCTCTGTAATCA---CAAAATGTACATATGAAATGAAATCTCTTCA 18793
QY 1851 AAAGTTTCAATTTACTTAAGTCCAGATCCATCCAGAGAGATCCACTTCAATGCCA 1910
DB 18794 GAAGTTTCAATTTACTTTGCTCCAGATCCAT---CAGAGATCACTCTGTATGCCA 18848
QY 1911 GTTATAGCTTATGGAATGATTTCTTCAATATATAGGCTTGAAGTGAATTAATCTCT 1970
DB 18849 GCTATAGCTTATGGAATATATTTCTTAGGTATATGAAGCTGAAGTGAATTAATGTTT 18908
QY 1971 TGATCCATTTTCTGCAAAATATGATGTTGTG 2000
DB 18909 TGATCCATGAGGCTGTGGAATGATGTTGTG 18938

RESULT 20
ABD32953
ID ABD32953 standard; DNA; 173564 BP.
XX
AC ABD32953;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD20-011.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
XX leukemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PE 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
XX 14-MAR-2003; 2003US-0038838.
XX 15-APR-2003; 2003US-00417375.
XX 13-JUN-2003; 2003US-00461862.
XX 15-SEP-2003; 2003US-0063431.
XX 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polymucleotides and polypeptides useful
XX for diagnosing, preventing or treating cancers, especially lymphoma and
XX leukemia, or in screening for agents that modulate cancer.
```

PS claim 16; seqid 688; 310bp; English.  
 XX The invention relates to an isolated nucleic acid comprising at least 10  
 CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
 CC in the specification, or its complement. The nucleic acid encode cancer-  
 CC associated proteins. Also included are an expression vector comprising  
 CC the isolated nucleic acid cited above, a host cell comprising the above  
 CC recombinant nucleic acid or expression vector, a microarray for detecting  
 CC a cancer-associated (CA) nucleic acid comprising at least one probe  
 CC comprising at least 10 contiguous nucleotides of any of the above-  
 CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
 CC an open reading frame of a CA sequence selected from any of the 95  
 CC polynucleotide sequences as mentioned in the specification, or its  
 CC complement), an isolated antibody, (or its antigen binding fragment) that  
 CC binds to the above polypeptide, a hybridoma that produces the above  
 CC monoclonal antibody, a pharmaceutical composition comprising the above  
 CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
 CC cells (comprising the antibody cited above, methods for diagnosing cancer  
 CC or for detecting the presence or absence of cancer cells in an  
 CC individual, a method for inhibiting growth of cancer cells in an  
 CC individual, a method for delivering a therapeutic agent to cancer cells  
 CC in an individual, an electronic library comprising the above  
 CC polynucleotide or polypeptide (or their fragments), methods of screening  
 CC for anticancer activity or for a bioactive agent capable of modulating  
 CC the activity of a CA protein (CAP), methods for detecting cancer  
 CC associated with expression of a polypeptide in a test cell sample, a  
 CC method for treating cancers and a method for inhibiting the expression of  
 CC CA gene in a cell. The composition and methods are useful for detecting,  
 CC diagnosing, preventing and treating cancers, especially lymphoma and  
 CC leukaemia. These may also be used in screening for agents that modulate  
 CC cancer. The present sequence is a human CAP genomic sequence. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_poc\_sequences  
 XX  
 XX  
 SQ Sequence 173564 BP; 48418 A; 33739 C; 37150 G; 54257 T; 0 U; 0 Other;

Query Match 22.9%; Score 457.4; DB 13; Length 173564;  
 Beest Local Similarity 73.9%; Pred. No. 1.8e-68;

Matches 749; Conservative 0; Mismatches 206; Indels 58; Gaps 11;

QY 974 TAGTACAGGATACCTTGAGATACCTGAGGTTGGTTCATACACCAATTAATACAA 1033  
 DB 77710 TAGTACAGGATACCTTGAGATACCTGAGGTTGGTTCATACACCAATTAATACAA 77769  
 QY 1034 ATATGCAAGAGTGATATCAATTAAGAGTCAACAGCTTTGGCTTCCAGTG 1093  
 DB 77770 AGGTGATTAAGAGTATGAAAT-----TTTGGTTTCCAGTG 77810  
 QY 1094 CATATTAAGTTTGGTATATACCTGAGTCTCTTAAGTGTGCAATAGTATATGTC 1153  
 DB 77811 CATATTAAGTTTGGTATATACCTGAGTCTCTTAAGTGTGCAATAGTATATGTC 77869  
 QY 1154 TAAAAAACAACATCTTAATTTTAAATGCTTTATTAATAAATGCTTAACATCAT 1213  
 DB 77870 TAAAAAACAACATCTTAATTTTAAATGCTTTATTAATAAATGCTTAACATCAT 77927  
 QY 1214 TGAGCATTCAGTGAAGTTGTAATCTTTTCTGCTGAGAGTCTT-----TTCTTAT 1264  
 DB 77928 TGAACCTTCAGTGAAGTGTACTTTTCTGCTGAGAGTCTTGGCTTGATGTTATAT 77987  
 QY 1265 TGATGATGATCGGGGTGAGGTGCTGAAGCTTAAGGTGCTGCTGAGCATTTCTTAA-- 1322  
 DB 77988 TGTGATCTTATCAGAGGTGCTGAGGTGCTGAGGTGCTGCTGAGCATTTTAAAG 78047  
 QY 1323 ---ACAAAGTGAAGATTCGAATATCAGTGAATCTTCTTCATGAAGATTTCTCT 1379  
 DB 78048 AAGACAAGATATGTTTACCATTTGATGATCTCTTCTTCTGACAGATTTCTCT 78107  
 QY 1380 AGTGTGTGATGCTTTTGTATGATTTTATGACAGTGAAGCTTTGTAATTTGA-- 1437  
 DB 78108 AGCATGTGTGCTGTTTGTATGATTTTCTTAAGTGAAGCTTTGTAATTTGAAGT 78167

QY 1438 TCAATCTCTCAAAACCTGCTGCTTAAACACCTAATTAATTAATTTCTGAATCC 1497  
 DB 78168 CAATCTCTCTCAAGACTCTCTGTTTATTTAGAGAAATTTAAATATTTCTAATATC 78227  
 QY 1498 ATTGTTGTCATTTCAACAAATTTTCAAGTGTCTTCAACAGAGATGATTCATCATTT 1557  
 DB 78228 TTTGCTGTCATTTCAACAAATTTTCAAGTGTCTTCAACAGAGATGATTCATCATTA 78287  
 QY 1558 CCTGAGATGGAATCTTTGCTGATCAATGAAGAAATTTCTGATGCTCAAGTTTATC 1617  
 DB 78288 AACCACTT---TCTCTGTTCAATGAAGAACTTCAACAGCTTCAAGTTTATC 78343  
 QY 1618 ATGAGATTCAGCAATATACATGATGCTTCAAGGCTTCACTTCACTTTTATTCAGTTCT 1677  
 DB 78344 ATGAGATTCAGCAATATACATGATGCTTCAAGGCTTCACTTCACTTTTATTCAGTTCT 78398  
 QY 1678 CTGCTGTTTCAACATCTGAGTCTTCTTCTGATGAAGTCTTGAACCTTCCAG 1737  
 DB 78399 CCTGCTGTTTCAACATCTGAGTCTTCTTCTGATGAAGTCTTGAAGCTTCCAG 78458  
 QY 1738 TCATCAGAGGATGGAATCACTTCTTCCAAATTCCTGTTAATTAATTTTGA-- 1795  
 DB 78459 TCATCAGAGGATGGAATCACTTCTTCCAAATTCCTGTTAATTAATTTTGA-- 1795  
 QY 1796 -CCTCCATGAATCAATGATGTTCTTAATGAGCCTGGAATGATGCTTCCAAAG 1854  
 DB 78519 TCTTCCATGAATCAATGATGTTCTTAATGAGCCTGGAATGATGCTTCCAAAG 78578  
 QY 1855 GTTTTCAATTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1914  
 DB 78579 GCTTCAATTTAATTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 78633  
 QY 1915 TGCCCTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1967  
 DB 78634 TAGCCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 78683

# RESULT 21

ACN44286/C  
 ID ACN44286 standard; DNA; 105219 BP.

XX ACN44286;  
 AC ACN44286;  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human genomic sequence hCG36856.  
 XX  
 KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003073826-A2.  
 XX  
 PD 12-SEP-2003.  
 XX  
 PF 28-FEB-2003; 2003WO-US006235.  
 XX  
 PR 01-MAR-2002; 2002US-00087192.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW;  
 XX  
 DR WPI; 2003-328604/31.  
 XX  
 PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 PT comprises a nucleotide sequence.  
 XX  
 PS Claim 1; SEQ ID NO 658; Opp; English.  
 XX  
 CC The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable



CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biobchip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent sequence to basic patent  
 CC US2002182586A1, for which no sequence data was published

XX Sequence 105219 BP; 30555 A; 20106 C; 20937 G; 33432 T; 0 U; 189 Other;

Query Match 22.7%; Score 454.8; DB 11; Length 105219;

Best Local Similarity 75.0%; Pred. No. 4.9e-68; Matches 780; Conservative 0; Mismatches 187; Indels 73; Gaps 14;

```

OY 977 TACAGGCAATCCTTGAGATAGTCTGGGTTGGTTCATACCAACCAATAATACAAATA 1036
DB 104524 TACAGGCAATCCTTGAGATAGTCTGGGTTGGTTCATACCAACCAACCAAT- 104466
OY 1037 TGCAAGAAATGATATACAAATAAAGTAGTACACAAAGCTTTGGCTCCCAAGTCAAT 1096
DB 104465 -----AATPAAATAGTACACAAATAATTTTGGTTTCCCAAGTCAAT 104425
OY 1097 ATAAAGTTTGTCTATATACATAGTCTGTGTAAAGTGCATATAGTATGTCTAA 1156
DB 104424 ATAAAGTTTGTCTATATACATAGTCTGTGTAAAGTGCATATAGTATGTCTAA 104365
OY 1157 AA---AAACACATACCTTAATTT-TAAATGCTTATTAATAAAA-ATGCTAACATC 1210
DB 104364 AAACCAATATATATATATATTAATAAATGTTCTATGCTAATAAATGCTAATATC 104305
OY 1211 ATTGAGCATTTAGAGATGTTGTAATCTTTTGTGTGTAAGAGTCTTTTCTATATGATA 1270
DB 104304 ATCTAGGCTTTAGAGATGTTGTAATCTTTTGTGTGTAAGAGTCTTTTCTATATGATA 104253
OY 1271 CTGATCGGGGTCAGGTCGTCGTAAGTCTAGAGTGTGCTGCTGCTTAA-----ACA 1325
DB 104252 -----GATGTAAGTGTGCTGTAAGTGTGCTGCTGCTGCTTAAATAAATAAGAA 104200
OY 1326 ACAGTGAAGATTTGCAATATATCAATGTAATCTTTTCAATGAAAGATTTCTCTAGTGTG 1385
DB 104199 GCAATGAATTTGCCAATCAATATGTAATCTTTTCAAGAGGATTTCTCTAGTGTG 104140
OY 1386 TGATGCTTTTGTATAGCATTTTATGCAAGTGAACCTTTTGAATAATGGA-TCATATC 1444
DB 104139 TGATGCTTTTGTATAGCATTTTATGCAAGTGAACCTTTTGAATAATGGAATC 104080
OY 1445 TCTCAAAACCTGCTGCTTTTAAACAACCTAAGTTAATATATATGCAATCATTGTTG 1504
DB 104079 TCTCGAATCCTGCACTGTTTATTTAATAGTTTATGTTATATCTTAATATCTTTGTTG 104020
OY 1505 TCATTTCAACAATTTTCAAGTGTCTTCAACAGAGTAGAATTCATCTCAATTTCTGAGA 1564
DB 104019 TCATTTGACAGGTTTACAGCTCTTCAACAGAGTAGAATTCATCTCAATTTCTGAGA 103960
OY 1565 TGAATCTTTGCTCATCATAGAGAATAATCTCTCATCTGTTCAAGTTTATCATGAGAT 1624
DB 103959 T-----TCTTTGTTCAATCCATAGAAACAATCTCTCATCTG-----TTAGTTTGAAGAT 103911
OY 1625 TGAGCAATATACAGTCAATGCTTCAAGGCTCATTCTTTAATTCAGATCTCTTGCGT 1684
DB 103910 TGAGCAATATACAGTCAATGCTTCAAGGCTCATTCTTTAATTCAGATCTCTTGCGT 103856
OY 1685 TTTCTACCAATCTGTGTTCTTCTCTCAATGAAGCTTGAACCTTCCCAAGTCAATCCA 1744
DB 103855 TTTCAACACACCTGCAATCTTCAATCACTGAAGCTTGAACCTTCAAGTCAATCCA 103796
OY 1745 TGAGGCTTGAATGCAATCTTCTCAATTCCTGTATATATATATATGACCT---CC 1800
DB 103795 AGAGATTTGAATCAATCTTCTCAATTCCTGTATATATGATATCTTACATCTCC 103736

```

```

OY 1801 CATGAATGATGATGTTCTTATAGGCACCTGGAATGTGAATCCTTTCCAAAAGGTTTC 1860
DB 103735 CATGAATGATGATGTTCTTATAGGCACCTGGAATGTGAATCCTTTCCAAAAGGTTTC 103677
OY 1861 AATTTACTTAGTCCAGATCCATCCATCCAGAGATCCACTTTCATATGCAATATAGCCT 1920
DB 103676 AATTTACTTAGTCCAGATCCATCCATCCAGAGATCCACTTTCATATGCAATATAGCCT 103622
OY 1921 TATGAATGATGATGTTCTTATAGGCACCTGGAATGTGAATCCTTTCCAAAAGGTTTC 1980
DB 103621 TATGAATGATGATGTTCTTATAGGCACCTGGAATGTGAATCCTTTCCAAAAGGTTTC 103562
OY 1981 TCTGCAAAATGATGTTGTG 2000
DB 103561 GCTGCAAAATGATGTTGTG 103542

RESULT 22
ABD32602
ID ABD32602 standard; DNA; 277616 BP.
XX
XX
AC ABD32602;
XX
XX 18-NOV-2004 (first entry)
DT
XX
DE Human cancer-associated genomic DNA HD7-221.
XX
XX Human; ds; cancer-associated protein; gene; cytosolic; cancer;
XX leukaemia; lymphoma; CAP.
XX
XX Homo sapiens.
XX
XX WO2004074320-A2.
XX
XX 02-SEP-2004.
XX
XX 17-FEB-2004; 2004WO-US004730.
XX
XX 14-FEB-2003; 2003US-00367094.
XX
XX 14-MAR-2003; 2003US-00388938.
XX
XX 15-APR-2003; 2003US-00417375.
XX
XX 13-JUN-2003; 2003US-00461862.
XX
XX 15-SEP-2003; 2003US-00663431.
XX
XX 15-DEC-2003; 2003US-00737318.
XX
XX (SAGE-) SAGES DISCOVERY INC.
XX
XX Morris DW, Morris DW, Malandro MS;
XX
XX MPI; 2004-652914/63.
XX
XX New isolated cancer-associated polynucleotides and polypeptides useful
XX for diagnosing, preventing or treating cancers, especially lymphoma and
XX leukemia, or in screening for agents that modulate cancer.
XX
XX claim 16; seqid 109; 310pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising at least 10
XX contiguous nucleotides of any of the 233 polynucleotide sequences given
XX in the specification, or its complement. The nucleic acids encode cancer-
XX associated proteins. Also included are an expression vector comprising
XX the isolated nucleic acid cited above, a host cell comprising the above
XX recombinant nucleic acid or expression vector, a microarray for detecting
XX a cancer-associated (CA) nucleic acid comprising at least one probe
XX comprising at least 10 contiguous nucleotides of any of the above-
XX mentioned nucleotide sequences, an isolated polypeptide (encoded within
XX an open reading frame of a CA sequence selected from any of the 95
XX polynucleotide sequences as mentioned in the specification, or its
XX complement), an isolated antibody, (or its antigen binding fragment) that
XX binds to the above polypeptide, a hybridoma that produces the above
XX monoclonal antibody, a pharmaceutical composition comprising the above
XX antibody and a pharmaceutical excipient, a kit for detecting cancer

```



PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0228287P.  
PR 01-SEP-2000; 2000US-0228343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235844P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241211P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249254P.  
PR 17-NOV-2000; 2000US-0249255P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251855P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM,  
XX WPI; 2001-502630/55.  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases.  
XX  
XX Disclosure; SEQ ID NO 4100; 986bp; English.  
XX  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
XX encoding a digestive system antigen of the invention  
XX  
XX Sequence 32190 BP; 10714 A; 6313 C; 5965 G; 9198 T; 0 U; 0 Other;

Query Match 22.7%; Score 454; DB 4; Length 32190;

Best Local Similarity 74.1%; Pred. No. 6,5e-68;

Matches 781; Conservative 0; Mismatches 215; Indels 58; Gaps 14;

QY 970 TAGATAGTACAGGCATACCTGGAGATACCTGCTGGCTTGGATTCATACCAACCATTAAT 1029  
DB 29296 TATTAAATACAGGTAAACCTTGAAGATATTCAGTTTGGTTCCAGACCATTCATTA 29355  
QY 1030 ACAATATGCAAGANAGTGAATATCACATTAAGTGAACACACAGCTTTGGCTTCCC 1089  
DB 29356 GCAG-----ATATCAAAATTAAGACGTTTACACAAACCTTTGT-TTCCC 29399  
QY 1090 AGTCATATTAAGTTTGGCTTAACTACACCTGTAAGTCTGTTAAGTGTGCAATAGTTTA 1149  
|||||

Db 29400 AGTGCATATTAACAT----TTATATATATTTGTAATCTAGTAAGTGTGTAATACCAATTA 29455  
QY 1150 TGTCT-AAAAAACACATACCTTAAT-TTAAAAACCTTATTAACATAAAAAATGCAACA 1207  
Db 29456 TGGCTGAAAAATATATATCAATTCAAAAATATTTTAAAGCTAAAAAATGCTAAAG 29515  
QY 1208 ATCATTTGAGCATTCAGTAGTGTATCTTTTGGCTGGAGAGGCTTTTCTTAATGA 1267  
Db 29516 TTCACTGAGCCCTTCGCAAGATGTGTCTTTTGGTGGAGAGGAGCTTACCTTATAT 29575  
QY 1268 TGAAGTAT-----CGGGGGTCAGGTGCTGAAGCTTGAAGTGGGTGGTGGCAGTTT 1316  
Db 29576 TGATGGCTGCTGATATGATCTGGGTAGTGGTGGCTCAAGGTGGGGGTGGCTGGCAACTT 29635  
QY 1317 CTTAAACG----ACGTGAAGTTCGCAATATAGTGTACTCTTCTTTCATGAAGATT 1372  
Db 29636 CTTAAATATGACATGAAGAGTTGGCTGCACAACTGACTTTCTTTTCACAAAGATT 29695  
QY 1373 TCTCTAGTGTGTAAGCTTTTGTATAGCATTTTATGACAGTATGACCTTTGAAAA 1432  
Db 29696 TATCTGTATGTGTGATGCTGTTTGTATGACATTTTACCAAGTGAAGCTTCTTCAAAA 29755  
QY 1433 TTGGATC-AATCCTCAAAACCCCTGCTGCTTTAAACAACCTAAGTTAATATATATCT 1491  
Db 29756 CTGGAGCTATCTCTTCAAAACCTGCTGCTTTATCACTAATTTATATATCTCT 29815  
QY 1492 GAATCATTTGTTGTCATTTCACAATTTTTCACAGTCTTTCACAGAGTATGATTCATC 1551  
Db 29816 GAATCTTTGTTGTCATTTCACAAGTGTTCACAGCATCTTTCACAAAGTATGATTCATC 29875  
QY 1552 TCATTTCTGAGATGGAATCTTGTCTCATCCATTAAGAAATTCCTCATCTGTTCAAGT 1611  
Db 29876 TCAAGAACCACTT---TGGTGTCTGTCCAGAAAGAACCACTCTCATTCACAGT 29931  
QY 1612 TTATCATGAGATGAGCAATACAGTCAATGCTTCAAGGCTCACTTCACTTTAATTC 1671  
Db 29932 TTTATCATGAGATGAGCAATACAGTCAATGCTTCAAGGCTCACTTCACTTTAATTC 29986  
QY 1672 AGTCTCTGCTGTTTCTACACATCTGTGTCTTCTCTCATTTGAAGCTTTGAACCTTC 1731  
Db 29987 AATTCCTGCTGTTTCTACACATCTGTGTCTTCTCTCATTTGAAGCTTTGAACCTTC 30046  
QY 1732 TCCAGTCAATGAGAGGTTGGAATCGACCTTCTCAAAATCCTGTTAATTTAATTT 1791  
Db 30047 TAAACTCATATGAGAAATGAAATGAATCTTCTTCCAACTCTGTATATGTTAATTT 30106  
QY 1792 TTGA---CCTCCATGATATGATGATGTTCTTAATGGACCTGGAATGGAATCTTTC 1848  
Db 30107 CTGACCTCTCTCCGTAATCAAAATGTTCTTAATGGCATCTAAGATGGAATCTTTC 30166  
QY 1849 CAAAAGTTTCAATTTACTTATGTCAGATTCATTCATCAAGATTCATTCATTC 1908  
Db 30167 CAGATGTTTAAATTTACTTGGCCAGCTCCAT-----CAGAGGACTCATATCATAT 30221  
QY 1909 CAGTTATACCTTANAGATGT--ATTTCTCAATTAAGGCTTGAAGTTGAATTC 1966  
Db 30222 CAGCATATACCTTANAGATGT--ATTTCTCAATTAAGGCTTGAAGTTGAATTC 30281  
QY 1967 TCTTGATCATTTTCTGCAAAATAGATGTG 2000  
Db 30282 TCCCTGACTCA-TGGCTGTGAATGATGTGTG 30314

RESULT 24  
ADCS5392  
ID ADCS5392 standard; DNA; 96587 BP.

XX AC ADCS5392;  
XX DT 01-JAN-2004 (first entry)  
XX DE Human Jak2 genomic sequence.

KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;  
KW secreted; transmembrane; intracellular; ds.  
OS Homo sapiens.  
PN WO2003045230-A2.  
PD 05-JUN-2003.  
PF 02-DEC-2002; 2002WO-US038582.  
PR 30-NOV-2001; 2001US-00997722.  
PA (SAGR-) SAGRES DISCOVERY.  
PI Morris DW, Engelhard EK;  
DR WPI; 2003-513603/48.  
PT New recombinant nucleic acid comprising a nucleotide sequence of any of  
PT the carcinoma-associated (CA) genes, useful for screening for drug  
PT candidates for diagnosing or treating carcinomas.  
PS Claim 1; SEQ ID NO 178; 983bp; English.  
XX  
XX The invention relates to a recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the fully defined carcinoma-  
CC associated (CA) genes from the 50 tables given in the specification. The  
CC CA proteins are secreted, transmembrane or intracellular proteins. The  
CC recombinant nucleic acids are useful for screening for drug candidates  
CC for diagnosing or treating carcinomas. Sequences given in ADCS5215-  
CC ADCS5514 represent CA genes of the invention.  
SQ Sequence 96587 BP; 28110 A; 17495 C; 18527 G; 32435 T; 0 U; 20 Other;  
Query Match 22.7%; Score 453; DB 10; Length 96587;  
Best Local Similarity 73.1%; Pred. No. 9.9e-68;  
Matches 764; Conservative 0; Mismatches 225; Indels 56; Gaps 12;  
QY 973 ATATGACAGGACCTTGGAGATACGTGGGTTGGTTCATACCAACCAATATACA 1032  
Db 51948 AAAATATGACACACCTCGAGATATATGGGTTGGTTCAGACTATCGAGTGAAGA 52007  
QY 1033 AATATGCAAGAGTGAATATCAATTAAGTATGACACAAAGCTTTTGGCTCCAGT 1092  
Db 52008 AATATG-----CAATTAAGCAAAATCACTGAAGTTTGAAGTTTCCAGT 52052  
QY 1093 GCATTAAGATTTGCTTATCTACCTGATGCTGCTGTTAAGTGAATAGTATATGT 1152  
Db 52053 GCATTAAGATTTGCTTATCTTATCTGATGCTGTTAAGTGAATATATATATGT 52112  
QY 1153 CTTAAAAAACACATACCTTAATTTTAAATGCTTATTAATTAATAAATGCTAACAATCAT 1212  
Db 52113 CCAAAAAGTACATACCTTAATTTTAAATTAATTTATTTATGTCAAAAGCACTCATATCAT 52172  
QY 1213 TTGAGCATTCAGTATGTTTAATCTTTTGTGCTGGAAGGCTTTTCTTATTTG----- 1266  
Db 52172 CTGACCTTCAGTATGATTAATCTTTTGTATGGAAGAGCTTGCCTGTATGTTGATG 52232  
QY 1267 ----ATGACGATGCGGGGGTCAAGTGAAGCTTAAGGTTGGCTGAGTTCTTAA 1322  
Db 52233 GCTTACTGCTGACAGGGGGTGGTGGCTGTAAGGTTGGGTGGCTATGCAATATCTTAA 52292  
QY 1323 ACA--ACAAGTGAAGTTGCAATATGATGATCTCTCTCTTTCATGAAGATTTCTCTTA 1380  
Db 52293 ATTAAGCATATGAAGGTTGTGATCATGATGATCTCTCTCTTCTGTAAGATGCTCTGTA 52352  
QY 1381 GTGTGTGATGCTTTTGAATGATATTTATGACATTAAGAACTTTTGAAGAAATGGA--T 1438  
Db 52353 GTATGCAATGCTTATTTGAATGATTTTACCGGCAATGAAGAACTTTTGAAGAAATGGAAGT 52412  
QY 1439 CAATCTCTCAAAACCTGCTGCTTTTAAACAACCTAAGTTAATATATTTCTGATCCA 1498

Db 52413 TGATTCACCAACCCCTGCCACTGCTT--TGATTAAAGTGAATGATTAATTAATCTT 52469  
Qy 1499 TTGTTGTCATTTCAACAATTTTTCACAGTGTCTTTCACAGAGTAGATTCATCTCATTTTC 1558  
Db 52470 TTGTTGTCATTTCAACAATTTTTCACAGAGTGTCTTTCACAGAGTAGATTCATCTCAAGA 52529  
Qy 1559 CTGAGATGGAATCTTTGCTCATCCATTAAGAGAAATTCCTCATCTGTTCAAGTTTATCA 1618  
Db 52530 ACCACTT---TCTTGGCTCATCCATTAAGAGAAATTCCTCATCTGTTCAAGTTTATCG 52585  
Qy 1619 TGAGATTCAGCAATACAGATCATGTCCTTCAGGCTTCACCTTCTTAAATCCAGTCTTC 1678  
Db 52586 TGAGATTCAGCAATACAGATCATGTCCTTCAGGCTTCACCTTCTTAAATTCAGTCTTC 52640  
Qy 1679 TTGCTGTTTCTACACATCTGTGTCTTCTTCTTCATGAAGTCTTGAACCTTCCAAAGT 1738  
Db 52641 TTGTTACTTTTACACATCTGTGTCTTCTTCTTCATGAAG-ATTAAATCCCTCAAGA 52699  
Qy 1739 CATCATGAGGAGTGAATGCATCTTCTTCCAAATTCCTGTAAATTTATATTTTGAAGCT 1798  
Db 52700 TACCATGAGAGCCAGAAATCACCTTCTTCCAAATTCCTGTAAATTTATATTTTGAAGCT 52759  
Qy 1799 CC---CATGAATCATGAAGTGTCTTAAATGAGCCTGGAATGGGAATCCTTCCAAAGG 1855  
Db 52760 CCTCTCATGAATCATGAAGTGTCTTAAATGAGCCTGGAATGGGAATCCTTCCAAAGT 52819  
Qy 1856 TTTTCAATTTTCTTACATCTGTGTCTTCTTCTTCATGAAGTCTTGAACCTTCCAAAGT 1915  
Db 52820 TTTTATTTTCTTACATCTGTGTCTTCTTCTTCATGAAGTCTTGAACCTTCCAAAGT 52870  
Qy 1916 AGCCTTAAAGAAATGATTTCTTCAATTAATTAAGGCTTGAAGTGAATTAATCTTGAATC 1975  
Db 52871 AACCTTAAAGAAATGATTTCTTCAATTAATTAAGGCTTGAAGTGAATTAATCTTGAATC 52928  
Qy 1976 CATTTTTCGCAAAATAGAGTTGTG 2000  
Db 52929 CATGGCTGCGAATGAGATGTTGTG 52953

RESULT 25  
ADA02912  
ID ADA02912 standard; DNA; 96589 BP.  
XX  
XX  
AC ADA02912;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human JAK2 carcinoma associated gene, SEQ ID NO:1430.  
XX  
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ds.  
XX  
OS Homo sapiens.  
OS  
PN WO2003057146-A2.  
XX  
PD 17-JUL-2003.  
XX  
PF 26-DEC-2002; 2002WO-US041414.  
XX  
PR 26-DEC-2001; 2001US-00035832.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW;  
XX  
DR WPI; 2003-587068/55.  
XX  
PT New recombinant nucleic acid encoding carcinoma associated protein,  
XX useful for preparing compositions for treating carcinomas.  
PS Claim 1; SEQ ID NO 1430; 245BP; English.

XX  
CC The invention relates to recombinant carcinoma associated (CA) nucleic  
CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or  
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed human CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcr\_sequences.

Sequence 96589 BP; 28111 A; 17494 C; 18529 G; 32435 T; 0 U; 20 Other;

Query Match 22.7%; Score 453; DB 9; Length 96589;

Best Local Similarity 73.1%; Pred. No. 9.9e-68;

Matches 764; Conservative 0; Mismatches 225; Indels 56; Gaps 12;

Qy 973 ATAGTACAGGACATCTTGGAGATACCTGTGGTGTGTTCCATACACCAATTAATCA 1032  
Db 51948 AAAATATGAGACACACCTGGAGATATGTGGTGTGTTCCAGATATCGAGTAAACA 52007  
Qy 1033 AATATGCAAGAAGTGGATATCAACAATAAGTAGACACACAAGCTTTTGGCTTCCAGT 1092  
Db 52008 AATATG-----CAATTAAGCAATATCAATGAAGTTTAGGTTTCCAGT 52052  
Qy 1093 GCATATTAAGTTTGTCTTAACTACACTGATGCTGTAAAGTGTGCAATAGTGTATGT 1152  
Db 52053 GCATATTAAGTTTGTCTTAACTATCTGTATGCTGTAAAGTGTGCAATATATATGT 52112  
Qy 1153 CTAATAAACAACATACCTTAATTTAAATGCTTATTAATTAATAAATGCTTAACATCAT 1212  
Db 52113 CCAAAAAAGTACATACCTTAATTTAAATCTTATTTGTCAAAAAGCATCATATCAT 52172  
Qy 1213 TTGACATTCAGTGTGTTGATCTTTTGTGCGTGAAGCTTCTTATTTG----- 1266  
Db 52173 CTGAGCTTCAGTGTGATATATCTTTTGTGAGAGAGTCTTGATGTGATG 52232  
Qy 1267 ---ATGACTGATCGGGGGTCAAGTGTGAAGCTTAGGGTGGCTGGCAGTTTCTTAA 1322  
Db 52233 GCTACTGATGACAGAGGGTGGTGGCTGCGTAAGTGTGGGTGGCTGAATATCTTAA 52292  
Qy 1323 ACA--ACAAGTAAGTGCATATATGATGACTCTTCTTCTTCAATAAAGATTCTCTTA 1380  
Db 52293 ATTAAGCAATGAAGTGTGCTGATGATGACTCTTCTTCTTGAAGATGTCTGTA 52352  
Qy 1381 GTGTGTGATGCTTTTGAATAGATTTTATGACAGATGAACCTTCTTGAATTTGA--T 1438  
Db 52353 GTATGCAATGCTATTTGATAGATTTTACCCGAGATGAACCTTCTTGAATAAGAGT 52412  
Qy 1439 CAATCTCTCAAAACCTGCTGTCTTAAACAATTAATTAATTAATTTCTGAATCA 1498  
Db 52413 TGATTCACCAACCTGCACTGCT--TGATTAAAGTGAATGATTAATTAATCTT 52469  
Qy 1499 TTGTTGTCATTTCAACAATTTTTCACAGTGTCTTTCACAGAGTAGATTCATCTCATTTTC 1558  
Db 52470 TTGTTGTCATTTCAACAATTTTTCACAGAGTGTCTTTCACAGAGTAGATTCATCTCAAGA 52529  
Qy 1559 CTGAGATGGAATCTTTGCTCATCCATTAAGAGAAATTCCTCATCTGTTCAAGTTTATCA 1618  
Db 52530 ACCACTT---TCTTGGCTCATCCATTAAGAGAAATTCCTCATCTGTTCAAGTTTATCG 52585







QY	1556	TTTTTCATTATTA	CTTAGTCAGATCCATCCAGAGAGATCCACTTTCAGTCCAGTTAT	1913
Db	52820	TTTTTTATTATTA	CTTTGGCTTAGATCTGT-----GAGAGGAAACACT-----GTGACAGCCAT	52870
QY	1916	AGCCTTAGGAAATG	TATTTCTTCATATATAAGGCTTGAAAGTTGAATTA	CTCTTGATC 1975
Db	52871	AACCTTACTTA	ATATGTGTTTCCCAATATATAGATTGAAGT--CAATTTGCTCTTATC	52928
QY	1976	CAITTTCTGC	AAATAGATGTTGTG 2000	
Db	52929	CATGGCTGC	AGATGATGTTGTG 52953	
RESULT 28				
ABL64107/c				
ID	ABL64107	standard; DNA; 14117 BP.		
XX	ABL64107;			
AC				
XX	15-MAY-2002	(first entry)		
DE		Breast cancer related gene sequence SEQ ID NO:2444.		
XX				
KW		Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;		
KW		stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;		
KW		cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;		
XX		gene; ds.		
OS		Homo sapiens.		
XX				
XX	WO200194629-A2.			
XX				
XX	13-DEC-2001.			
XX				
PF	30-MAY-2001;	2001WO-US010838.		
PR	05-JUN-2000;	2000US-0209473P.		
PR	05-JUN-2000;	2000US-0209531P.		
PR	18-SEP-2000;	2000US-0233133P.		
PR	18-SEP-2000;	2000US-0233617P.		
PR	20-SEP-2000;	2000US-0234009P.		
PR	20-SEP-2000;	2000US-0234034P.		
PR	20-SEP-2000;	2000US-0234052P.		
PR	22-SEP-2000;	2000US-0234509P.		
PR	22-SEP-2000;	2000US-0234567P.		
PR	25-SEP-2000;	2000US-0234923P.		
PR	25-SEP-2000;	2000US-0234924P.		
PR	25-SEP-2000;	2000US-0235077P.		
PR	25-SEP-2000;	2000US-0235082P.		
PR	25-SEP-2000;	2000US-0235134P.		
PR	25-SEP-2000;	2000US-0235280P.		
PR	26-SEP-2000;	2000US-0235637P.		
PR	26-SEP-2000;	2000US-0235638P.		
PR	27-SEP-2000;	2000US-0235711P.		
PR	27-SEP-2000;	2000US-0235720P.		
PR	27-SEP-2000;	2000US-0235840P.		
PR	27-SEP-2000;	2000US-0235863P.		
PR	28-SEP-2000;	2000US-0236028P.		
PR	28-SEP-2000;	2000US-0236032P.		
PR	28-SEP-2000;	2000US-0236033P.		
PR	28-SEP-2000;	2000US-0236034P.		
PR	28-SEP-2000;	2000US-0236109P.		
PR	28-SEP-2000;	2000US-0236111P.		
PR	29-SEP-2000;	2000US-0236842P.		
PR	29-SEP-2000;	2000US-0236891P.		
PR	02-OCT-2000;	2000US-0237172P.		
PR	02-OCT-2000;	2000US-0237173P.		
PR	02-OCT-2000;	2000US-0237278P.		
PR	02-OCT-2000;	2000US-0237294P.		
PR	02-OCT-2000;	2000US-0237295P.		
PR	02-OCT-2000;	2000US-0237316P.		
PR	03-OCT-2000;	2000US-0237425P.		

PR	03-OCT-2000;	2000US-0237598P.	
PR	03-OCT-2000;	2000US-0237604P.	
PR	03-OCT-2000;	2000US-0237606P.	
PR	03-OCT-2000;	2000US-0237608P.	
PR	01-NOV-2000;	2000US-0244867P.	
PR	01-NOV-2000;	2000US-0245084P.	
PA	(AVAL-)	AVALON PHARM.	
PI	Young PR,	Augustus M,	Carter KC,
PI	Soppet DR,	Weaver Z,	Ebner R,
XX			Endress G,
XX			Horrigan S;
XX	WPI:	2002-188264/24.	
PT	Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.		
PT	Claim 1; SEQ ID NO 244; 44pp; English.		
XX	The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in AB161664 to AB170110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms's tumour		
XX	Sequence 14117 BP; 3568 A; 3182 C; 3484 G; 3883 T; 0 U; 0 Other;		
XX	Query Match		
XX	Best Local Similarity 72.6%; Score 451.6; DB 6; Length 14117;		
XX	Matches 766; Conservative 0; Mismatches 224; Indels 60; Gaps 12;		
QY	972	GATAGTACAGGCATACCTTGGAGTACTGTGGGTTTGCTTCATACCAACCAATATATAC	10311
DB	1108	GAAATCAACGGCATATACCTCAGAGATGTGCAATGTTCAAGTTCTAACCACCAATTAAGC	10499
QY	1032	AAATATGCAAGATGTGATATCAATTAAGTGTGACACAAAGTCTTTGGCTCCAG	10931
DB	1048	AAATGTA-----ACATTAAGTGTGACACAAACTTTTGGTTTCTG	10004
QY	1092	TGCATATAAAGTTTGGTTATATCTATACACTGTAGTCTGTTAAGTGCATATAGTTATG	11531
DB	1003	TGTATGTAAACATTATATTTAACACTATATAGTATTAAGTGAACATATAGATTATG	944
QY	1152	TCTTAAAAA---ACACATACCTTAATTTTAAAAAGCTTTATTAATCTAAAAAGTCTAACA	1207
DB	943	GCCCTAAAAACATGTATATTAATCTTAATTTAAAAAGTCTTTCTGTCTAAAAATCTAAG	884
QY	1208	ATCATTTGAGCATTCAGTGTGTGTATCTTTTGTGTGGAGAGGTCTTTCTTATATGA	1267
DB	883	ATCATTCGAGCTTCAGCAAGTTATTAATCTTTTGTGTGGAGAGGTCTTGTCCCTAAATGT	824
QY	1268	TGACTGAT-----CGGGGTCAAGTGTGAAAGCTTAGGGGTGCTGTGACGTT	1316
DB	823	TGATGGCTGAGGACTGATCAGGGGTGATGATTTGCTGCAAGTTAAGTGTGGCTGTGCAATTT	764
QY	1317	CTTAAA-----ACAAAGTGAAGTTGCATATATCAATTTACTTTCTTTCAATTAAGAT	1371
DB	763	CTGAAATATGAACTACATCGTAAAGTTTCCGCAATCAAGTTACTTTCTTTCAATTAAGAT	704
QY	1372	TTCTCTAGTGTGTGAGCTTTTATTAATGAGATTATTAAGACAGTATGAACTTTCTTGAA	14311

```
Db      703  TTCCCTGATGATGATGATGCTGTTGATAGCATTTTAA---ACAGTAAGAAATTCCTTCAAA 647
Oy      1432 A-TTGATGATCAATCCCTCAAAACCCCTGCTCTGCTTTAACAACAACTAGATTAATATATATTC 1490
Db      646  ACTAGGAGCAGTCCCTCTCAAAACCCCTGCACTGCTTATTAATTAATGATGATATTC 587
Oy      1491  TGAATCCATTTGTTGATTTCAATTTTCAACAAATTTTCAACAGTCTCTTACCAAGAGATGATTCAT 1550
Db      586  CAATTCCTTTGTCATTTGATGATGTTTCAACAGATCTTCAACAGAGATGATTCAT 527
Oy      1551  CTCATTTCTGATGATGATCTTTGCTCATTCATAGAAGAAATTCCTCATCTGTTCAAG 1610
Db      526  CTCAGAAACCACTT---TCTTTGTTCAATGATAGAAGAGCTTCATCCGTTCC--G 473
Oy      1611  TTTTATCATGAGATTTGACGACATACGATGATGCTTTGAGGCTCACTTCATTTAATTC 1670
Db      472  TTTTGCCATACATGAGCAACATTCAGTCAATCTTCAGGC-----TCAATTCATATT 419
Oy      1671  CAGTTCTCTGCTGTTTCTTACCAATCTGTGCTCCCTTCCTCATTTGAAGTGAACCT 1730
Db      418  TAATTCCTGTGCTGTTTCCAAACACATCTATAGTTACTTCCTCTGCTGAAGTCTTAAGCC 359
Oy      1731  CTCCAAGTCATCCATGAGGCTTGAATGCACTTTCCAAATTCCTGTTAATTAATTAT 1790
Db      358  CTCCAATCATATCAATGAGAGTTGGAATCAACTTCTTCCAACTCCGTTAAAGTGAT 299
Oy      1791  TTTGACCTCCCAATGATCATGATGTTCTTAATGGACCTGGATGATGATGCTTTTCA 1850
Db      298  TTTGACCTCTCTCTTGAATCA---CAATATTTGACATATAGATATGATATCCCTTCA 242
Oy      1851  AAAGTTTTCATTTACTTACTAGTCAGATCCATCCATCCAGAGATCCACTTCAATGCCA 1910
Db      241  GAAAGTTTTCATTTACTTACTTGGCCAGATCCAT----CAGAGGAATCA-TCTGATAGGA 188
Oy      1911  GTTATAGCTTATGGAATGATTTCTTCAATATATAGGCTTGAAGTTGAATTTACTCT 1970
Db      187  GCTATAGCCTTATGAAATATATTTCTTAGTATATAGACTTGAAGTCAAAATTAATTT 128
Oy      1971  TGATCCATTTTCGCAAAATAGATGTTG 2000
Db      127  TGATCCATGGGCTGTGGAATGATGTTG 98

RESULT 29
AAL05284
ID      AAL05284 standard; DNA; 4503 BP.
XX
XX      AAL05284;
AC
XX
XX      21-NOV-2001 (first entry)
DT
XX
XX      Human reproductive system related antigen DNA SEQ ID NO: 7972.
DE
XX
XX      Human; reproductive system related antigen; reproductive system disorder;
KW      cancer; gene therapy; ds.
XX
XX
OS      Homo sapiens.
XX
XX      WO200155320-A2.
PN
XX
XX      02-AUG-2001.
PD
XX
XX      17-JAN-2001; 2001WO-US001339.
PF
XX
XX      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205515P.
```

```
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      11-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226686P.
PR      22-AUG-2000; 2000US-0227182P.
PR      23-AUG-2000; 2000US-0227182P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231242P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      14-SEP-2000; 2000US-0233065P.
PR      14-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      29-SEP-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241221P.
```

XX	PR	20-OCT-2000;	2000US-0241785P.
XX	PR	20-OCT-2000;	2000US-0241786P.
XX	PR	20-OCT-2000;	2000US-0241787P.
XX	PR	20-OCT-2000;	2000US-0241808P.
XX	PR	20-OCT-2000;	2000US-0241809P.
XX	PR	20-OCT-2000;	2000US-0241826P.
XX	PR	01-NOV-2000;	2000US-0244617P.
XX	PR	08-NOV-2000;	2000US-0246474P.
XX	PR	08-NOV-2000;	2000US-0246475P.
XX	PR	08-NOV-2000;	2000US-0246476P.
XX	PR	08-NOV-2000;	2000US-0246477P.
XX	PR	08-NOV-2000;	2000US-0246478P.
XX	PR	08-NOV-2000;	2000US-0246522P.
XX	PR	08-NOV-2000;	2000US-0246524P.
XX	PR	08-NOV-2000;	2000US-0246525P.
XX	PR	08-NOV-2000;	2000US-0246526P.
XX	PR	08-NOV-2000;	2000US-0246527P.
XX	PR	08-NOV-2000;	2000US-0246528P.
XX	PR	08-NOV-2000;	2000US-0246532P.
XX	PR	08-NOV-2000;	2000US-0246609P.
XX	PR	08-NOV-2000;	2000US-0246610P.
XX	PR	08-NOV-2000;	2000US-0246611P.
XX	PR	08-NOV-2000;	2000US-0246613P.
XX	PR	17-NOV-2000;	2000US-0249207P.
XX	PR	17-NOV-2000;	2000US-0249208P.
XX	PR	17-NOV-2000;	2000US-0249209P.
XX	PR	17-NOV-2000;	2000US-0249210P.
XX	PR	17-NOV-2000;	2000US-0249211P.
XX	PR	17-NOV-2000;	2000US-0249212P.
XX	PR	17-NOV-2000;	2000US-0249213P.
XX	PR	17-NOV-2000;	2000US-0249214P.
XX	PR	17-NOV-2000;	2000US-0249215P.
XX	PR	17-NOV-2000;	2000US-0249216P.
XX	PR	17-NOV-2000;	2000US-0249217P.
XX	PR	17-NOV-2000;	2000US-0249218P.
XX	PR	17-NOV-2000;	2000US-0249244P.
XX	PR	17-NOV-2000;	2000US-0249245P.
XX	PR	17-NOV-2000;	2000US-0249246P.
XX	PR	17-NOV-2000;	2000US-0249254P.
XX	PR	17-NOV-2000;	2000US-0249255P.
XX	PR	17-NOV-2000;	2000US-0249297P.
XX	PR	17-NOV-2000;	2000US-0249299P.
XX	PR	17-NOV-2000;	2000US-0249300P.
XX	PR	01-DEC-2000;	2000US-0250160P.
XX	PR	01-DEC-2000;	2000US-0250391P.
XX	PR	05-DEC-2000;	2000US-0251030P.
XX	PR	05-DEC-2000;	2000US-0251988P.
XX	PR	05-DEC-2000;	2000US-0256719P.
XX	PR	06-DEC-2000;	2000US-0251479P.
XX	PR	08-DEC-2000;	2000US-0251856P.
XX	PR	08-DEC-2000;	2000US-0251868P.
XX	PR	08-DEC-2000;	2000US-0251869P.
XX	PR	08-DEC-2000;	2000US-0251989P.
XX	PR	08-DEC-2000;	2000US-0251990P.
XX	PR	11-DEC-2000;	2000US-0254097P.
XX	PR	05-JAN-2001;	2001US-0259678P.
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	F1	Rosen CA, Barash SC, Ruben SM;	
XX	DR	WPI; 2001-465570/50.	
XX	PT	Isolated nucleic acid molecule encoding a reproductive system antigen is	
XX	PT	used in preventing, treating or ameliorating a medical condition.	
XX	PS	Disclosure; SEQ ID NO 7972; 1297bp + Sequence Listing; English.	
XX	CC	The present invention provides the protein and coding sequences of a	
XX	CC	number of human reproductive system related antigens. These can be used	
XX	CC	in the prevention and treatment of reproductive system disorders,	
XX	CC	including cancer. The present sequence is a genomic sequence encoding a	
XX	CC	protein of the invention	

Seq	Sequence	4503 BP	1360 A	837 C	792 G	1514 T	0 U	0 Other
Query Match	22.4%	Score	447.8	DB	4	Length	4503	
Best Local Similarity	71.3%	Pred.	No. 6,86-67					
Matches	756	Conservative	0	Mismatches	242	Indels	63	Gaps 10
QY	970	TGATATGACGCGCATCTTGGAGATACGTGGGTTTGGTCCATACACACACATATAT	1029					
DB	99	TATGATATACAGGATACCTTTAGAGATGGTGCATATTTGGATTGACAGCACAGCAAAAA	158					
QY	1030	ACAAATATGCAAGAGTGGATATACAAATAAGAGATACACAAAGCTTTGGCTTCCC	1089					
DB	159	GCAATATCTG-----CAATACGCAAGTCACACAAAGTTTGGTTTCCC	203					
QY	1090	AGTCATATTAAGATTTTGGCTTATACACATCTAGTCTGTATAGTGCATATGTGTTA	1149					
DB	204	ACGCGATATGGAAGTATGTTTACGCTATCTAGTCTAGTATAGTGTGAAGAAAGCATTA	263					
QY	1150	TGTC-----TAAAAAACACATACCTTAATTT-TAAAAAGCTTTATTACTAAA	1197					
DB	264	TTTCTTTTTTTTTTTTTAAAGGACATACCTTATTTAGAAACGCTTACTCTAAA	323					
QY	1198	AATGCTAACATCATTTGAGATTCAGTGAATGTATATCTTTTGGCTGGGAAGTCTT	1257					
DB	324	AATGTAAACATCATCTGCACCTTCACACAGGTCTGAATCTTTTGGCCAGTGGAGGTCCT	383					
QY	1258	TTCTTTA-----TTGATGACCTGATCGGGGGTCAAGTGCCTAAGCTTAAGGTGGCTG	1307					
DB	384	GCTTCATATGTGATGGCTGCTGACTGATAGGTGTGTATTACTGAAAGTTGGAGATGCATC	443					
QY	1308	TGGCAGTTCTTTAAACA-----ACAGTGAAGATTTGCATATACGTGACTCTTCCTTC	1362					
DB	444	TGGCATTTCTTAATAATAGAAAGCATATGAATTTGGCAATCATGTTGATCTTCCTTC	503					
QY	1363	ATGAAGAATTTCTCTTAGTGTGATGCTTTTGTATAGACATTTTATGCAACGTAGAAT	1422					
DB	504	ACGAAGAATTTCCCTGATGACATGTAGTGTGTTCATAGACATTTTACCACAGTAGAAC	563					
QY	1423	TCTTTGAAAAATTTGA-TCAATCCCTCAAAOCCGTCTGTCTTAAACAACCTAATTAAT	1481					
DB	564	TCTTCAAAATTTGAGTCAAGTCTCTCAATCTGTGCACAGCTTTGCACTTAAGTTT	623					
QY	1482	ATAATATTTTGAATTCATATGTTGTTCATTTTCAACAATTTTCAAGTCTTCACAGGAGT	1541					
DB	624	GTCATATATCTATATCTTCATATGATATTTTCAACATATGTTCAATACTTCATCATGAT	683					
QY	1542	AGATTCATCTCATTTCTCGATGATGAACTTTGGCTCATGCATTAAGAAATTCCTCAT	1601					
DB	684	AGATTCATCTCAAGAAACACTT-----TCTTGGCTCATGCATTAAGAAATTAATCTCTGT	739					
QY	1602	CTGTTCAGTTTATCATGAGATTTGCGCAATATACATCATGTCTTCAGGCTCATCTTCAC	1661					
DB	740	CCATTCCAATTTTATCATGAGATTTGCGCAATATCAAGATCAATCTTCAGG-----CTCCAC	794					
QY	1662	TTTTTAATTTCCAGTCTCTCTGCTGTTTTCACACATCTGTGGTCTCTCTCATTTGAAT	1721					
DB	795	TTTCAATTTCTGGTCTCTCTGTGTTTTCACACATCTTCAGTCACTTCCTCTCTCAATAAGT	854					
QY	1722	CTTGAACCTCTCAAGTCATCCATGAGGT-----TGAATGCACTTCTTCCAAATTC	1775					
DB	855	CTTGAACCTCTCAAGTCATCCATGAGGTCAAAAAATGATTCATTTTCTTAAATCTCC	914					
QY	1776	TGTTAATATTTATATTTTGAATCTCCATGATCATGAATGTTCTTAATGGCACTGGAAT	1835					
DB	915	TGTTAATATTTATTTTGAATCTCTCTCTGATGATCATGATATCTCTTAATGCACTTCGAAT	974					
QY	1836	GGTGAATCTTCCAAAGGTTTTCATTTACTATAGTCCAGATTCATCCATCAGAGAT	1895					
DB	975	GATTAATTTTATTTTATATGTTTACATTTTATTTTGGCCAGATTCAT-----CAGAGAA	1029					
QY	1896	CCACTTTCATATCCAGTTATATAGCTTATGGAATGATATTTCTTCAATATATAGGCTTGAA	1955					
DB	1030	TCACTTACCTTGGGAGGCTATAGCTTATGGAAGTATATTTCTTAAGTATTAAGCTTGAA	1089					



```

ID      ADV77908 standard; DNA; 175590 BP.
AC      ADV77908;
XX      24-FEB-2005 (first entry)
XX      Human BAC clone RP11-114017 on chromosome 10 genomic DNA.
DE      Choline transporter; neuromuscular disorder;
XX      autonomic nervous system disorder; central nervous system disorder;
XX      Parkinson's disease; Huntington's chorea; genetic disorder;
XX      Alzheimer's disease; degeneration; neurological disease; schizophrenia;
XX      psychiatric disease; myasthenia gravis; immune disorder; gene therapy;
XX      cms-gen.; neurotropic; neuroprotective; antiparkinsonian; anticonvulsant;
XX      muscular-gen.; neuroleptic; DNA purification; BAC;
XX      bacterial artificial chromosome; RP11-114017; chromosome 10; gene; ds.
OS      Homo sapiens.
XX      US2004248838-A1.
XX      09-DEC-2004.
XX      01-DEC-2003; 2003US-00724806.
XX      23-JUL-2001; 2001US-00911077.
XX      (UYVA-) UNIV VANDERBILT.
XX      Blakely RD, Apparsundaram S, Ferguson S;
XX      WPI; 2005-020586/02.
XX      GENBANK; AC009933.
XX      New human high affinity choline transporter (CHT) cDNA, useful for
XX      treating a neuromuscular, autonomic or central nervous system disorder,
XX      including Parkinson's disease, Huntington's disease, Alzheimer's,
XX      schizophrenia.
XX      Example 1; SEQ ID NO 13; 73pp; English.
XX      The present invention provides polynucleotides encoding novel high
XX      affinity choline transporters (CHTs), methods for their use in screening
XX      and therapy. The invention is useful for treating neuromuscular
XX      disorders, autonomic or central nervous system disorders such as
XX      Parkinson's disease, Huntington's disease, Alzheimer's disease,
XX      schizophrenia, dysautonomia and myasthenia gravis (MG). The invention is
XX      also useful in gene therapy. The present sequence is human BAC (bacterial
XX      artificial chromosome) clone RP11-114017 on chromosome 10 genomic DNA.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from the US
XX      patent office at seqdata.uspto.gov/sequence.html?DocID=US20040248838.
XX      Sequence 175590 BP; 55029 A; 33099 C; 32863 G; 53896 T; 0 U; 703 Other;
XX      Query Match 22.3%; Score 446.8; DB 14; Length 175590;
XX      Best Local Similarity 73.3%; Pred. No. 1,1e-66;
XX      Matches 771; Conservative 0; Mismatches 217; Indels 64; Gaps 13;
XX      970 TAGATAGTACAGGCGATACCTTGGAAGTACTGTGGGTTTGCTTCATACACCAATTAAT 1029
XX      133582 TACAAAATATAGAGTAGCTTGGAAGTATTTGGGGGTTTGCTTCACAGACCTCAATA-- 133639
XX      1030 ACAAAATATGCAAGAGTGGATATACAAATTAAGTGTACACACAAGCTTTTGGCTTCC 1089
XX      133640 -----AAGCGAATGTCACAAATTAAGTGTACACACAATTTTGGTTTCC 133686
XX      1090 AGTGCAATTAAGAGTTTGTCTTATATCTAGTACTGTCTGTTAAGTGCATAGTGTTA 1149
XX      133687 AGTGCAATACAGAGTTGTGCTTCACTATATCTGTATATTTAAGTGTGAAGTGCAATTC 133746
XX      1150 TGTCT-AAAAAACAATCACTTAATTTTAAATGCTTTATTTATCTAAAAATGCTTAACA 1207
XX      |||

```

Db	133747	GGTTTGGAAATTACATACCCTAGTTAAAAAAATCCTTATGTGCAAAAAATGCTAATG	133806
Oy	1208	ATCATTTGAGCATTCAGTGAAGTTGTAATCTTTTGTCTGTGTGAAGAGTCTTTTCTTATTTGA	1267
Db	133807	ATCATCTGAGGCTTTAGTAGAGTCCAAATCTCTTGTGCAAGTGAAGGATCTTACCTCAGTGT	133866
Oy	1268	TGACTGAT-----GGGGGGTCAAGTGTCAAGCTTAAGGGTGGCTGTGGGAGTTT	1316
Db	133867	TAAATGGCTGCTGACTGATCAGGGGTGGTGTGTCTCAAGTTGGGGTGGCTGTGGGAAATTT	133926
Oy	1317	CT-----TAAACAACAGTGAAGAATGTGCATATCAGTTGACTCTTCTCTTCATGAAGAAT	1371
Db	133927	CTGAANAATAGACAACAATGAAGTTATTCGACATTGACGTGACCTCTCTTGGGAAATAT	133986
Oy	1372	TTCTCTCTAGTGTGATGCTTTTGTGATAGCATTTTATGACAGATGAACTTCTTTGAAA	1431
Db	133987	TTTTATGTGGCATGTGATATATGTTTGAATGACATTTTA---CCGATAGAACTTCTTAA	134043
Oy	1432	ATTGG-ATCATATCCCTGTAACCCCTGCTCTGTCTTTAACAAGCTAAGTTAATTAATATTC	1490
Db	134044	ATTGGAAATCAGCTCTTCCAAATCCATACGTCTAATTATACAGTAAGTTATATTAATATTC	134103
Oy	1491	TGAATTCATGTGTGTCATTTTCAACAATTTTCAACAGTGTCTTGCACAGAGATGATTCAT	1550
Db	134104	TTAAAT-TTTGTGTGTTATTTTAAACAATGTTCAAGAAATCTTCAACAGGAATTAATTTCAAT	134162
Oy	1551	CTCATTTCTGTGATGAGAAATCTTTGCTCATCCATAGAGAAGAAATTCCTCATCTGTTCAAG	1610
Db	134163	CTCATGAAGCAACACTCA-----TTTGCTCATCATTAAGAAGCATCTCTCATTTGTTCAAG	134218
Oy	1611	TTTTATCATGAGATTTGCAGCAATACAGTCATGTCTTCAGGCTCACTTCTTTAATTC	1670
Db	134219	TTTTATCAGAGACTGCACAATTCAGTCACATCTTCAGG-----CTCACCTTCTAATTC	134273
Oy	1671	CAGTTCCTGTCGTGTTCTTACCAACATCTGTGGTTCCTTCCTCATTTGAAGCTTGAACCT	1730
Db	134274	TAGTTCCTGTCGTGTTCTTACCAATATCTG-----CTCCTGTCTAAGTCTTGAACCT	134326
Oy	1731	CTCCAAAGTCATTCATGAGGTTTGAATCGACTCTTCCAAATTCCTGTATATATTTATAT	1790
Db	134327	CTCAAAAGTCATTCATGAGGTTTGAATCGAATTTCTTCCAAATTCCTGTATAGTTGATAT	134386
Oy	1791	TTTGAAC--CTCCATGATCATGTAATGTTCTTAAATGCACTGGAATGTGTAATCTTT	1847
Db	134387	TTTATACCTCTCTCATGATGACGAATATCTCTTAACACATCTGAAGATGCAAAATCTTT	134446
Oy	1848	CCAAAAGGTTTTCATTTACTTATGTCCAGATCCATCCATCAGAGAGATCCACTTTCAATG	1907
Db	134447	CCAAAAGGCTTTCAACTTACTTGTCCAGCTTCAT-----CAGAGAAATTAACAATCTTATG	134501
Oy	1908	CCAGTTATAGCCTTATGGAATGTATTTCTTCAATATATAGGCTTGAAGTTGAATTTACT	1967
Db	134502	GCA--ACTAGCCCTACAAAATGTATTTTCTTAAGTGTATTAACCTTGAAAGTCAAAATTTACT	134559
Oy	1968	CCTTGATTCATTTTCTGCAAAATAGATGTGT 1999	
Db	134560	CATTGATTCATGAGTACTACAGATGATGTGTG 134591	
RESULT 32			
ABQ88179			
ID	ABQ88179	standard, cDNA, 160771 BP.	
AC	ABQ88179;		
XX	18-SEP-2002	(first entry)	
DT			
XX			
DE	Human osteoblast differentiation related cDNA SEQ ID NO 86.		
XX	Human: osteoblast; stem cell differentiation; bone tissue deposition;		
KM	osteoporosis; osteopathic; ss.		
XX			
OS	Homo sapiens		

XX WO200250301-A2.  
 XX 27-JUN-2002.  
 XX 18-DEC-2001; 2001WO-US048276.  
 XX 18-DEC-2000; 2000US-0255882P.  
 XX 24-APR-2001; 2001US-0285691P.  
 XX (GENE-) GENE LOGIC INC.  
 XX (PROC) PROCTER & GAMBLE CO.  
 XX J. D., Axelrod DW., Cook JS., Jaiswal N., Einstein R., Houghton A;  
 XX Mertz L;  
 XX WPI; 2002-557663/59.  
 XX Use of genes and their expression profiles associated with osteoblast  
 XX differentiation for screening modulators bone formation, for diagnosing  
 XX or treating e.g. osteoporosis, or as markers for the differentiation  
 XX process.  
 XX Claim 1; SEQ ID NO 86; 78bp + Sequence Listing; English.  
 XX The invention relates to genes and their expression profiles are used  
 XX for: (a) screening modulators of precursor stem cell differentiation into  
 XX osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
 XX deposition of bone tissue, abnormal rate of osteoblast formation or  
 XX osteoporosis; or (c) treating or monitoring treatment of the conditions  
 XX cited in (b), or monitoring the progression of bone tissue deposition.  
 XX Specific conditions include postmenopausal osteoporosis, glucocorticoid  
 XX induced abnormalities in bone formation or bone loss, conditions that  
 XX involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
 XX skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
 XX or fibrous dysplasia. The present sequence is that of an osteoblast  
 XX differentiation associated CDNA marker of the invention. Note: The  
 XX sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 160771 BP; 44776 A; 35622 C; 34804 G; 45569 T; 0 U; 0 Other;  
 XX  
 XX Query Match 22.2%; Score 444; DB 6; Length 160771;  
 XX Best Local Similarity 73.2%; Pred. No. 3.4e-66;  
 XX Matches 770; Conservative 0; Mismatches 215; Indels 67; Gaps 13;  
 XX  
 XX 977 TACAGGCATACCTTGGAGATACCTGCTGGTTGCTTCATACCAACAATAATATACAAATA 1036  
 XX |||||  
 XX 67072 TACAGGCATACCTTGGAGATATGTAGATTAGTCCAGCCAGTGAAGACACATA 67131  
 XX |||||  
 XX 1037 TGCAGAGATGATATCACAAATTAAGTGTACACAAAGCTTTTGGCTTCCAGTGCAT 1096  
 XX |||||  
 XX 67132 G-----CTCAATGAATCAAGTCCACATTTTATTTATTTCCAGTGCAT 67176  
 XX |||||  
 XX 1097 ATAAAGTTTGGCTTATACACCTGTAGCTGTGAAGTGCAGAAATGTTATGCTAA 1156  
 XX |||||  
 XX 67177 ATAAAGTTATATTTACCTATACATAGCTATTAATTAAGTGAAGTGAATATGCTAAT 67236  
 XX |||||  
 XX 1157 AAAAAG-----ACATACCTTAATTTTAAATGCTTTATTAATAAATGCTAAACATCA 1211  
 XX |||||  
 XX 67237 AAAAAGCATATCATCTTATTTTAAAAATATTTATTTATGCTAAAAATGCTAATGATCA 67296  
 XX |||||  
 XX 1212 TTTGAGCATTCAGTGAAGTGTAAATCTTTTGTCTGTGGAAGTCTTTTCTTATGTAGAC 1271  
 XX |||||  
 XX 67297 TCTGAGCCTTGTGGTAGTAGTATCTTTTGTCTGTGAGAGGCTTGCTGTGATGTGAT 67356  
 XX |||||  
 XX 1272 TGAAT-----CGGGGCTAGAGTGTGAGGCTTAGAGGCTGTGGCAATTTCTTA 1320  
 XX |||||  
 XX 67357 GGCTGTGACTTATATCAGGAGAGTGTGTGGAAGGTTGGGGTGTGGGGCAATTTCTTA 67416  
 XX |||||  
 XX 1321 AA-----ACAAAGTGAAGATTGCATATCATGTTGACTTCTTTCATGAAGATTTCCT 1375  
 XX |||||

DB 67417 AAATPAGCAACAATGAAGTTCATCTGATGATTAATCTTGTCTTCAAGAAATTTCT 67476  
 |||||  
 QY 1376 CTTAGTGTGATGCTTTTGTATAGCATTTTATGACAGTAGAACTTTTGAATAATG 1435  
 |||||  
 DB 67477 CTGAGCATTAAGATGCTTTTGTAGCAACATTTTACCACAGTAGAATCTTTCAAAATG 67536  
 |||||  
 QY 1436 GATCAATCCTCTCAAAACCTGCTCTGTTTAAACAAGTAAATATATATATTTCTGAAT 1495  
 |||||  
 DB 67537 GAGTCA-GTACTCAAAACCTGCTGCTG-----TCAACTAAGTTTATGTAATCTTAAAT 67589  
 |||||  
 QY 1496 CCATTGTGTATTTCAACAATTTTACAGATGCTTCCACGAGATAGATTCATCTCAT 1555  
 |||||  
 DB 67590 CTTTGTGTGATTTTCAACAATTTTACAGATGCTTCCACGAGATAGATTCATCTCAT 1555  
 |||||  
 QY 1556 TTTCTGAGATGGAATCTTTGCTCATCTCATTAAGAAAT-----TCTCATCTGTTCAAG 1610  
 |||||  
 DB 67646 TCAAGGAACCACTCATCTCATCTCATCTCATCAAGAAAGCACTCACTCTCTGTTAAG 67705  
 |||||  
 QY 1611 TTTTATCATGATTTGACGAATATACATGATGCTTTCAGGCTCACTTCACTTTAATTC 1670  
 |||||  
 DB 67706 TTTGATCATGATGATCAGAAATTCAGTCCATCTTCAAG-----TCTTCACTTTAATTC 67760  
 |||||  
 QY 1671 CAGTCTCTGCTGTTTCTTACCAATCTGCTGCTTCTTCTTCTTCAAGTCTTGAACCT 1730  
 |||||  
 DB 67761 TAGTCTCTGTTTATTTTCCATCAATGTCAGATGATCTTCTTCACTTCAAGTCTTGAACCT 67820  
 |||||  
 QY 1731 CTCCAGTCAATCATGAGGTTTGAATGATGATCTTCTTCAAAATCTGTTAATATTAAT 1790  
 |||||  
 DB 67821 TTTCAAGTTATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 67880  
 |||||  
 QY 1791 TTTGA-----CTTCCATGATCATGATGATGATGATGATGATGATGATGATGATGAT 1847  
 |||||  
 DB 67881 TTTGACCTTCTTCCATGATCATGATGATGATGATGATGATGATGATGATGATGAT 67940  
 |||||  
 QY 1848 CCAAAAGTTTCAATTTACTTATGTCAGATGATGATGATGATGATGATGATGATGATGAT 1907  
 |||||  
 DB 67941 CCAAGAAATTTTCCATTTCACTTGTGCTGAGATGCA-----CAGGGGACTCACTAATTTATG 67995  
 |||||  
 QY 1908 CCAAGTATAGCTTATGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1967  
 |||||  
 DB 67996 GCACTTACAGGTTTCAAAATATATTTCTTAAATAT--GACTTGAATGCAAAATTAAT 68053  
 |||||  
 QY 1968 CCTGATCCATTTTCTGCAAAATAGATGTTGT 1999  
 |||||  
 DB 68054 GCTCATTCACAGCAGCAGAAATAGATGTTAT 68085  
 |||||  
 XX  
 XX RESULT 33  
 XX AAK90584  
 XX ID AAK90584 standard; DNA; 4320 BP.  
 XX  
 XX AAK90584;  
 XX  
 XX 05-NOV-2001 (first entry)  
 XX  
 XX Human digestive system antigen genomic sequence SEQ ID NO: 4160.  
 XX  
 XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
 XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 XX digestive system disorder; Meckel's diverticulum; ds.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200155314-A2.  
 XX  
 XX 02-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US001324.  
 XX  
 XX 31-JAN-2000; 2000US-0179065P.  
 XX  
 XX 04-FEB-2000; 2000US-0180628P.  
 XX  
 XX 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0198674P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-020515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225113P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226688P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 05-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229501P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 21-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234977P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235835P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 XX

PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251983P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI, 2001-502630/55.  
 XX  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases.  
 XX  
 PS Disclosure; SEQ ID NO 4160; 986bp; English.  
 XX



CC The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention  
XX

Sequence 4320 BP; 1282 A; 853 C; 822 G; 1363 T; 0 U; 0 Other;

Query Match 22.2%; Score 443.6; DB 4; Length 4320;

Best Local Similarity 74.2%; Pred. No. 3.5e-66;

Matches 739; Conservative 0; Mismatches 204; Indels 53; Gaps 12;

```
QY 1040 AAGAGTGGATATCATCAATAAAGTGCACAGTCTTTGGCTCCAGTGCATATA 1099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 372 ATGAGCGCATATCATCAATGACAAATCATATGATTTTGGTTTCCAGTGCATATA 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1100 AAAGTTTGCCTATACACTAGTCTGTTAAGTGCATATGTTATGCTTAAAAA 1159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 432 AAAGTTATGTTACACTACAGTGGTCTATTAAATGTCAGAACATTATGCTGAGAA 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1160 A--ACACATACCTTAATTTAAATGCTTTATCTATAAAAAATGTAACATCATTTGA 1216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 492 ATATATACACACCTTAATTTAAATGTTGCTT---AAAAGATGCTAACCAATAGTCAGA 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1217 GCATTGAGTGGTGTGAATCTTTTGGTGGAGGCTTTTCTTATGATGACTG--- 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 549 GCTTTATGAGAGTCTGCT-ATCTTTTGGTGGAGGCTTTTGGCTGATGTTGATGCTG 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1274 -----ATCGGGGTCAGGTCGTCGAGCTTAGGGTGGCTGAGCTTTCTTA----- 1320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 608 GTGACTGATTTAGGGTGGTGCAGTGGTGAAGTTGTATGCTGTGGAATTTCTTAAAAAT 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1321 -----AAACAACTGTAAGATTCGAATATCATGTTGACTCTTCTTTCATGAAAAA 1370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 668 ACAATTTCTTAAACACATGAAATGACATTCGACATGATGACTCTTCTTATGAAAAA 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1371 TTTCTCTAGTGTGATGCTTTTGTATGAGCATTTTATGACAGTAAACTCTTTGAA 1430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 728 TTTCTGTAGCTGCAGATGATTTTGTATGATGACTTTTGTCTGAGATGAACTTTCTTCAA 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1431 AATTGGA-TCAATCTCTCAAAACCTGCTCTGCTTTTACACCTTAAGTTAATATAAT 1489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 788 AATGGAGACAGTCCCTTAAACCCGCTGCTACTTATCACTAATGTTAATATGTTGG 847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1490 CTGAATTCATGTTGTCATTTCAACAATTTTCACTAGTCTTCCACAGAGATGATTCGA 1549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 848 CTAAATCCCTTGTGTATCATTTCAACAATGTTCAAGACATCTTAGCCAAAGTAAACCCA 907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1550 TCTCATTTCTGAGATGAAATCTTTGCTCATCCATAGAAAGAAATCTCATCTGTTCAA 1609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 908 TCTCAAGAACTACTT---TCTTGTCTGCTCATTA--AAGGACTCTCCCAACATTAA 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1610 GTTTTATCATGAGATGACAGCAATACATCATGCTTTCAGGCTCACTTC-ACCTTAAAT 1668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 962 GTTTTATCATGAAATGACAGCAATTCAGTCACATCTTCAGGCTCACTTCTATTAAGAAAT 1021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1669 TCCAGTCTCTTGTGCTTCTTCAACACATCTGTGCTCTTC---TCCATGAAAGTCTT 1724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1022 TCTAGTCTCTTGTCTTCTTCAATCATCTGAGATTACTTATCTTTTCACTTTAAAGTCTT 1081
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1725 GAACCTTCCAGTCAATCATGAGGATGGAATGCACTTCTTCAAAATCTGTTAAAT 1784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1082 GAACCCCTCAAGTCAATCATGAGATGGAATTAATCTTCCAAACCTCTTCAATG 1141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1785 TTAATATTTGACCTCCCATGAATCATGAATGTTCTTAATGAGCACTGAAATGCTGAATCC 1844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1142 TGAATATTTGATTTCCATGAAATCATGAATATTTCTTAATGCACTGAAATGCTGAATCC 1201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1845 TTTCCAAAAGTTTCAATTTACTTAATGTCAGATTCATTCATCAAGAGATTCACCTTTCA 1904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1202 TTT---ACAGGTTTCAATTTACTTGTCCCAAGCTCAT-----CAGAGAAATCATATCT 1253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 1905 ATGCCAGTTATAGCCTTATGGAATGTATTTCTCAATATATAGCCTGAAAGTTGAAT 1964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1254 GTGGACGCTTACAGCCTTACAAATGATATTTCTGATATATAGCTTATATGTTGAAATA 1313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1965 ACTCTTATTCATTTTCTGCAAAATATAGATGTTGG 2000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1314 ACTCTGATTCATGAGCTGCGAGAACACGCTGTGG 1349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 34

AAK90585 standard; DNA; 4320 BP.

AAK90585;

05-NOV-2001 (first entry)

Human digestive system antigen genomic sequence SEQ ID NO: 4161.

Human; digestive system antigen; gene therapy; cancer; appendicitis;

ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;

digestive system disorder; Meckel's diverticulum; ds.

Homo sapiens.

WO20015314-A2.

17-JAN-2001; 2001WO-US001334.

31-JAN-2000; 2000US-0179065P.

04-FEB-2000; 2000US-0180628P.

24-FEB-2000; 2000US-0184664P.

02-MAR-2000; 2000US-0186350P.

16-MAR-2000; 2000US-0189874P.

17-MAR-2000; 2000US-0190076P.

18-APR-2000; 2000US-0198123P.

19-MAY-2000; 2000US-0205515P.

07-JUN-2000; 2000US-0209467P.

28-JUN-2000; 2000US-0214886P.

30-JUN-2000; 2000US-0215135P.

07-JUL-2000; 2000US-0216647P.

11-JUL-2000; 2000US-0216880P.

11-JUL-2000; 2000US-0217487P.

11-JUL-2000; 2000US-0217496P.

14-JUL-2000; 2000US-0218290P.

26-JUL-2000; 2000US-0220963P.

26-JUL-2000; 2000US-0220964P.

14-AUG-2000; 2000US-0224518P.

14-AUG-2000; 2000US-0224519P.

14-AUG-2000; 2000US-0225213P.

14-AUG-2000; 2000US-0225214P.

14-AUG-2000; 2000US-0225266P.

14-AUG-2000; 2000US-0225267P.

14-AUG-2000; 2000US-0225268P.

14-AUG-2000; 2000US-0225270P.

14-AUG-2000; 2000US-0225447P.

14-AUG-2000; 2000US-0225477P.

14-AUG-2000; 2000US-0225757P.

14-AUG-2000; 2000US-0225758P.

14-AUG-2000; 2000US-0225759P.

18-AUG-2000; 2000US-0226279P.

22-AUG-2000; 2000US-0226681P.

22-AUG-2000; 2000US-0226686P.

22-AUG-2000; 2000US-0227182P.

23-AUG-2000; 2000US-0227009P.

30-AUG-2000; 2000US-0228924P.

01-SEP-2000; 2000US-0229287P.

01-SEP-2000; 2000US-0229343P.

01-SEP-2000; 2000US-0229344P.

01-SEP-2000; 2000US-0229345P.

05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234987P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239355P.  
PR 13-OCT-2000; 2000US-0239373P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.

PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250310P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251038P.  
PR 05-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI, 2001-502630/55.  
XX  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases.  
XX  
XX Disclosure; SEQ ID NO 4161; 986pp; English.

XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
XX encoding a digestive system antigen of the invention

XX  
XX Sequence 4320 BP; 1282 A; 854 C; 822 G; 1362 T; 0 U; 0 Other;  
Query Match 22.2%; Score 443.6; DB 4; Length 4320;  
Best Local Similarity 74.2%; Pred. No. 3.5e-66;  
Matches 739; Conservative 0; Mismatches 204; Indels 53; Gaps 12;

QY 1040 AAGAGTGATATCATTAAGAGTACACAAAGTCTTTGGCTTCCAGTCAATATA 1099  
DB 372 ATGAAGCGAGATATCAATTAAGAAATCAATGATTTTGGTTCCAGTCAATATA 431  
QY 1100 AAGTTTGGCTTATCTACACTGATGCTGTTAAGTGCATAGTATGCTTAATAA 1159  
DB 432 AAGTTATGTTACCTACCTGATGCTGTTAATTAATGTCAGAACATTAATGTCAG 491  
QY 1160 A---ACACATACCTTAATTTTAAATGCTTATTAATAAAATGCTAACATCAATTTGA 1216  
DB 492 ATATATACACACCTTAATTTTAAATGTTGCT---AAAAGTGTAAACATTAATGTCAG 548  
QY 1217 GCATTCAGTGAAGTTATATCTTTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1273  
DB 549 GCTTTTATGATGCT-ATCTTTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 607  
QY 1274 -----ATCGGGGGTCAGGCTGAAGCTTAAAGGTCGTCGTCGTCGTCGTCGTCGTCG 1320  
DB 608 GTGACTGATTAAGGGTGGTGAAGTGTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 667  
QY 1321 -----AAACAACGTGAAGTTCGAATATCATGTTACTCTTCTTCATGAAGA 1370  
DB 668 ACAATTTCTTAACAACATGAATGCCACATTAATGATGTCGTCGTCGTCGTCGTCGTCG 727

```

OY 1371 TTTCTCTAGTGTGTGATGCTTTTGTATGACATTTTATGACAGTAGACTCTTTGAA 1430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 728 TTTCTCTAGTGTGTGATGCTTTTGTATGACATTTTGTCTGAGATAGACTCTTTCAA 787
OY 1431 AATTGGA-TCAATCTCTCAAAACCTGCTGCTTTTAAACAACCTAAGTTAATATATAT 1489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 788 AATGGAGACAGTCTCTTAAACCTGCTGCTAATTATCAAGTTAATATATATAT 847
OY 1490 CTGAATCAATGTTGTCTATTTTCAACAATTTTCAAGTGTCTTCAACAGAGATGTTCCA 1549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 848 CTAATCTCTTGTATGATTTTCAACAATTTTCAAGAGATGTTCAAGAGATGAGACCA 907
OY 1550 TCTCATTTCTGAGATGAAATCTTGTCTCATCTCAAGAGAAATTTCTCATCTGTGAA 1609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 908 TCTCAAGAACTACTT-----TCTTGTGCTGCTCATT--AAGGAGCTCTCAACCTTAA 961
OY 1610 GTTTTATCATGATGATGAGCAATACAGTCAATGCTTCAAGGCTCACTTC-ACITTTAAT 1668
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 962 GTTTTATCATGAAATGAGCAATTCAGTCAATCTTCAAGGCTCAAGCTTCAATTAAGAT 1021
OY 1669 TCCAGTTCTCTGCTGCTTTTCTACCAATGCTGCTTCCCTCC-----TCCATTGAAGTCTT 1724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1022 TCTAGTTCTCTGCTGCTTTTCTACCAATGCTGCTTCAATCTTCTTCAATTAAGTCTT 1081
OY 1725 GAACCTCTCAAGTCAATCATGAGGTTGGAATTCACATCTTCTCAAAATTCCTGTTAATAT 1784
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1082 GAACCTCTCAAGTCAATCATGAGGTTGGAATTCACATCTTCTTCAAAATTCCTGTTAATAT 1141
OY 1785 TTAATATTTGACCTCCCATGAATCATGAATGTTCTTAATGCACTGGAATGCTGAATTC 1844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1142 TGAATATTTGATCTTCCATGAATCATGAATATTTCTTAATGAGATCATGAATGCTGAATTC 1201
OY 1845 TTTCCAAAAGTTTTCATTTTCTAATGTCAGATCCATCCAGAGATCCACTTCA 1904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1202 TTT---ACAGGTTTCAATTTACTTGGCCAGCTTCAT-----CAGAGAAATCACTAATCT 1253
OY 1905 ATGCCAGTTATAGCCTTATGATGATGATTTCTTCAATATATAGGCTTGAAGTTGAATAT 1964
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1254 GTGGAGCTCAAGCCTTACAAATATGATTTCTCGATTATATAGACTTGAATGTTGAATA 1313
OY 1965 ACTCCTGATTCATTTTCTGCAAAATATGATGTTGTG 2000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1314 ACTCCTGATTCATGAGGCTGAGAAACAGCTGTTGTG 1349

```

```

PA (DEAN/) DEAN N M.
PA (BEN/) BENNETT C F.
XX Bhanot S, Dobie KW, Freier SM, Dean NM, Bennett CF;
PI WPI; 2005-521414/53.
DR
XX New antisense compound targeted to a nucleic acid molecule encoding
XX glucocorticoid receptor, useful for modulating the expression of
XX glucocorticoid receptor, or for treating or diagnosing, e.g. diabetes or
XX obesity.
PS Example 12; SEQ ID NO 25; 129bp; English.
XX
XX The invention relates to an antisense compound 13-80 nucleobases in
XX length targeted to a nucleic acid molecule encoding glucocorticoid
XX receptor, where the compound is complementary to the nucleic acid
XX molecule encoding glucocorticoid receptor, and where the compound
XX inhibits the expression of glucocorticoid receptor mRNA. Also described:
XX (1) a method of inhibiting the expression of glucocorticoid receptor in a
XX cell or tissue; (2) a method of treating an animal having a disease or
XX condition associated with glucocorticoid receptor; (3) a method of
XX decreasing blood glucose levels in an animal; and (4) a method of
XX preventing or delaying the onset of an increase in blood lipid levels in
XX an animal. The compounds, compositions, and methods are useful for
XX modulating the expression of glucocorticoid receptor. They are also
XX useful for the diagnosis and treatment of diseases and conditions
XX associated with glucocorticoid receptor, e.g. diabetes (Type 2), obesity,
XX metabolic syndrome X, hyperglycemia, or hyperlipidemia. The present
XX sequence represents a human glucocorticoid receptor nucleotide sequence,
XX which is used in an example from the present invention.
SQ Sequence 32767 BP; 9265 A; 5801 C; 6236 G; 11465 T; 0 U; 0 Other;

Query Match      22.1%; Score 442.6; DB 14; Length 32767;
Best Local Similarity 71.5%; Pred. No. 5,66-66; Indels 64; Gaps 9;
Matches 734; Conservative 0; Mismatches 229;

OY 974 TAGTACAGGATACCTTGGAGATGATGCTGTTGGTTTCCATACCAACCAATTAATACAA 1033
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30604 TAAATACAGGATACCTTGGAGATGATGCTGTTGGTTTCCATACCAACCAATTAATACAA 30554
OY 1034 ATATGCAAGAGTGAATATCACATTAAGTGAATGATGATGATGATGATGATGATGATG 1093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30553 -----AAGTGAATATGCAACAAGCAAGTCAAGATTTTGTGTTCCACATCA 30503
OY 1094 CATATTAAGTTTGTGTTATATCTATACCTGATGCTGTTAAGTGAATGATGATGATGATG 1153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30502 CATATTAAGTTTGTGTTATATCTATACCTGATGCTGTTAAGTGAATGATGATGATGATG 30443
OY 1154 TAAATAAACAATACCTTATTTTAAATGCTTTTATTAATAAATATGCTTAACATCATAT 1213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30442 TTTAAATAATGATAT--TTACCTTAATCTTAAATATATGCTTTAAATATGCTTAACATCAT 30386
OY 1214 TGAGCATTCAGTGAATGTTAATCTTTTGTGTTGAGAGTCT-----TTTCTTAA 1263
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30385 TGAACCTTCAGTGAATGTTAATCTTTTGTGTTGAGAGTCTCTTCAATGTTGATGATG 30326
OY 1264 TTGATGATGATGAGGAGGATGAGTGTGAAGCTTGAAGTGTGAGGAGGAGGAGGAGGAGG 1323
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30325 CTGCTGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 30267
OY 1324 CAACAGTGAAGATGCAATATATGATGATGATGATGATGATGATGATGATGATGATGATG 1383
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30266 -----GTAGACAACAATGAAATTTTCACTTTTCTTTTCAATAAATTTCTTTTGAAGCA 30214
OY 1384 TGTGATGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1442
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30213 CGAGATGCTGTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 30154
OY 1443 CTTCTCAAAACCTGCTGCTGCTTTTAAACAACCTAAGTTAATATATTTTGAATCAATGCT 1502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30153 CCACTCAAAACATGCTGCTGCTTTTAAACAATTAAGTTAATATTTTAAATTCCTTTGT 30094

```

Oy	1503	TGTCATTTCAACAATTTTCCACAGTCTTCCACGAGAGTATGATTCATCTCATTTCTCGA	1562
Db	30093	TGTTATTTTCAACAAGTTTCCACAGATTTTCCACGAGAGTATGATTCATCTCAAGAAAAA	30034
Oy	1563	GATGGAATCTTTGGCTCATCCATAAGAGAAATTCCTATCTGTGTCAAGTTTATCATGAG	1622
Db	30033	ATTATATTTGCTCATTTTGTCCATTAAGAGCAATCTCTATCATTCATCAAGTTTTCATGAG	29974
Oy	1623	ATTGACGCAATACAGTCATGTCTTCAGGCTCATCTTCATCTTTAATTCAGTTCTCTGCG	1682
Db	29973	ATTGACGCAATTCGGTCACATTTTCGGG-----CTCACCTTCATATTCATGATCTCTTGG	29919
Oy	1683	TGTTTCTTACCAATCTGTGGTTCCTTCTCCTCATTGAAGTCTTGAACCTCTCAAGTCATC	1742
Db	29918	TGTTTCTTACCAATCTGTGAGTCTTCTTAACTGAAGTGTGAAACCCCTCAATTCATC	29858
Oy	1743	CATGAGGTTGGATTCGACTTCTCCAAATTCCTGTTAATTTATTTTGAACCTCC--C	1801
Db	29858	TATATCTGTTGGATTCAACTTCTTTCAAACGTCTATTAAGTTGATCTTTGACCTCCTC	29799
Oy	1802	ATGATCATGATATGTTCTTATATGCACTGTGAATGTGAATCCCTTCCAAAGATTTTCA	1861
Db	29798	CTGGGAATCACAAATTTCTTAATAGCACTGAATGCGCAATCTTTTCAGAGGTTTTTG	29739
Oy	1862	ATTTACTTAGTCCAGATTCATCCATCCAGAGATTCACATTTCATGCGCAGTTATAGCTT	1921
Db	29738	ATTTACTTTGGCCAGGTCCA-----CCAGAGCAATACGTCTATAGCAGCTATAGCTT	29688
Oy	1922	ATGGAATGTATTTCTTCAATPATTAAGGCTTGAAGTGAATAATCTCTTGATCCATTTT	1981
Db	29683	ACAAAAATGTACT-----GATTTGAAACTTGAATAATCACTCCCTGATTCATGGG	29633
Oy	1982	CTGCAAA	1988
Db	29636	CTACAAA	29630

RESULT	36
ABX08336_06	
Continuation (7 of 17) of ABX08336 from base 600001 (Human phosphodiesterase 4D (PDE4D	
WP Sequence split into 17 fragments	LOCUS ABX08336 Accession ABX08336
WP Fragment Name	Begin End
WP ABX08336_00	1 110000
WP ABX08336_01	100001 210000
WP ABX08336_02	200001 310000
WP ABX08336_03	300001 410000
WP ABX08336_04	400001 510000
WP ABX08336_05	500001 610000
WP ABX08336_06	600001 710000
WP ABX08336_07	700001 810000
WP ABX08336_08	800001 910000
WP ABX08336_09	900001 1010000
WP ABX08336_10	1000001 1110000
WP ABX08336_11	1100001 1210000
WP ABX08336_12	1200001 1310000
WP ABX08336_13	1300001 1410000
WP ABX08336_14	1400001 1510000
WP ABX08336_15	1500001 1610000
WP ABX08336_16	1600001 1691080
 Query Match 22.1%; Score 441.6; DB 6; Length 110000; Best Local Similarity 71.7%; Pred. No.8.6e-66; Indels 66; Gaps 11; Matches 755; Conservative 0; Mismatches 232;	
OY 973 ATAGTACAGGCATCTGTTGGAGATCTGTGGGTTTGATTCCATACCAACAATAATATACA	1032
Dd 51179 ATATACCAGACACTACTCGGAGATATTGTGGGTTTCAGATCAGGCCACTGCAATTAAGCA	51238
OY 1033 AATATGCAAGAAGTGGATATTCACAAATTAAGTAGTACACACAAAGTCTTTGGCTTCCAGT	1092
Dd 51239 A-----ATATTCACAAATTAAGCAAGTACACAACTCTTTTGGCTTCCAGT	51283

QY 1093 GGATTTAAAGTTTGGCTTATCTACACGTGAGTCTGTTAAGTGCAGATAGGTTATGT 1152  
 Db 51284 GGAATTAAGACTTATGTTTACCTATATTTGAGTGTGTA---TAGATTTCTCTAAG 51339  
 QY 1153 CTAAAAAACAACATACCTTAATTTTAAATGCTTTATTACT-----AAAAATGCTAAC 1207  
 Db 51340 AAAACAATGTACATGCTTAAGTAAGAAATTAATTAATGCTAATAAAAAAATGCTAC 51399  
 QY 1208 ATCATTTGAGCATCTCAGTGAATTGTAATCTTTTGTGCTGGGAAGTCTTTCTTATTTGA 1267  
 Db 51400 ACCATTAGAGCCTTCAGCAAGTCAGAAACATTTTGTGCTGGAGAGTCTTCCCTCAATGT 51459  
 QY 1268 TGACTGAT-----CGGGGGTCAGGTGCTGAAGCTTGGGGTGGCGTGGGCACTT 1316  
 Db 51460 TCATGATCTGCTGATCTGATCAAGGTGGTGGTGTCTGAAGATGGGGTGGCGTGAACATTT 51519  
 QY 1317 CT-----TAAACAACAGTGAAGATTTGCAATATACATTTGACTCTTCTTTCATGAAGAT 1371  
 Db 51520 CTTAAGATTAACCTACAGTGAAGAAATTTGCCACATCCATTAATTCCTCTTTCATGAACAT 51579  
 QY 1372 TTCTCTTATGTGTGATGCTTTTGTGATGACATTTTATGCAAGTGAACCTTCTTGA 1431  
 Db 51580 TTCTCTGATGACATGATGATGTTTGAACAGCATTTTACCATTGATGAACCTTCTTCAA 51639  
 QY 1432 ATTTGA-TCAATCCCTCTCAACCCCTGCTGCTTTAACAACCTAATTAATAATATTC 1490  
 Db 51640 ATTTGAGTCAATCCCTTAAAGCTGCTGCTGTTATCTCAATTTTAATAATAATTT 51699  
 QY 1491 TGAATCCATTTGTTGTCATTTCAACAATTTTCAACAGTCTTCAACAGAGTAGATTTCCAT 1550  
 Db 51700 TAAATCTTTTGTGTCATTTCAACAATGTTCAACAGATCTTTAACAAGAGTAGATTTCTGT 51759  
 QY 1551 CTCAATTCCTGATGAAATCTTTGCTCATCAATAAGAAATTTCTCATCTGTTCAAG 1610  
 Db 51760 CTCAGAAGAACCAATTTCTTTGCTCATC-----CAAGAAACAACCTCATTTCTTCAAG 51814  
 QY 1611 TTTTATCATGATATTCAGACCATTAAGTATGTTTCAAGGCTCATCTTCACTTTAATTC 1670  
 Db 51815 TTTTATCATTAATATTCAGTAATTTTCATTAACAATTCAG3-----TTCAATTTCTAATTC 51869  
 QY 1671 CAGTTCTCTTGCTGTTTCTACACATCTGTGTTCTTCTCTCCATGGAATCTTGAACCT 1730  
 Db 51870 TAGTTCTCTTGAATTTCCAGACATCTGCACTTCTTCTCA-----CTTGAACCC 51922  
 QY 1731 CTCCAGTCAATCATGAGGGTTGAATCGACTTTTCCAAATTCCTGTTAATATTAAT 1790  
 Db 51923 CTCAGATCAATCAGAGGGTTGAATCAATTTCTTCCAACTCTGTTAATGTTATAT 51982  
 QY 1791 TTTGACCT---CCCATGATCATGATGTTCTTAATGGAACCTGGAATGCTGAATCCTTT 1847  
 Db 51983 TTTGACCTATGCTTAATGATCAAAATATTTTAATGAGATCTAAGATGCTGAATCTGT 52042  
 QY 1848 CCAAAAGGTTTCAATTAATTAAGTCCAGATCCATCCATCCAGAGATCCAATTTCAATG 1907  
 Db 52043 CCAGAAATTTTTCGACTTAATCTTTGCCCTGATCAT-----CAGAGAAATTAAGTGTATC 52097  
 QY 1908 CCAATTAATAGCCTTAATGAATGATTTCTTCAATAATAAGGCTTGAAGTTGAATTAAT 1967  
 Db 52098 ATAGGATAGCCCTTAATGAATGATTTCTTAATATATAAGGCTTAATAAGTTGAATGACT 52157  
 QY 1968 CCTTATCATTTTCTGCAAAATAGATGTTGG 2000  
 Db 52158 CCTTGAACATGGGCTTCAGATGAATGTTTGG 52190

WP	Fragment Name	Begin	End	Accession
ADJ25985_02		200001	310000	
ADJ25985_01		100001	210000	
ADJ25985_00		1	110000	
ADJ25985_06				
ADJ25985_37				

WP	ADJ25985_03	300001	410000	
WP	ADJ25985_04	400001	510000	
WP	ADJ25985_05	500001	610000	
WP	ADJ25985_06	600001	710000	
WP	ADJ25985_07	700001	810000	
WP	ADJ25985_08	800001	910000	
WP	ADJ25985_09	900001	1010000	
WP	ADJ25985_10	1000001	1110000	
WP	ADJ25985_11	1100001	1210000	
WP	ADJ25985_12	1200001	1310000	
WP	ADJ25985_13	1300001	1410000	
WP	ADJ25985_14	1400001	1510000	
WP	ADJ25985_15	1500001	1610000	
WP	ADJ25985_16	1600001	1691139	
Query Match 22.1%; Score 441.6; DB 12; Length 110000;				
Best Local Similarity 71.7%; Pred. No. 8.6e-66; Mismatches 232; Indels 66; Gaps 11;				
Matches 755; Conservative 0;				
Qy	973 ATAGTACAGGACATACCTTGAGAGATCTGTGGTTGGTTCCATACACCAATATATACA	1032		
Db	51238 ATATACGAGACATACCTTGAGAGATATTTGGTTCCAGATCCAGGCCACTGCAATTAAGCA	51237		
Qy	1033 AATATGCAAGAGTGAATATCAATTAAGTGCACACAGTCTTTGGCTTCCAGT	1092		
Db	51298 A-----ATATCAAAATTAAGCAACATATATCTTTGGCTTCCAGT	51342		
Qy	1093 GCATATTAAGTTGCTTATCTACCTGCTGCTGTTAAGTGCATATAGTTATGT	1152		
Db	51343 GCATATTAAGTTGCTTATCTACCTGCTGCTGTTAAGTGCATATAGTTATGT	51398		
Qy	1153 CTAAAAAACACATACCTTAATTTTAAATGCTTATTAAT	1207		
Db	51399 AAAACATGATATCTGCTTAAGTAAATATCTTAATGCTTAAAAAATGCTAAACA	51458		
Qy	1208 ATCATTTGAGCATGAGTGTGTAATCTTTTGTGTGGAAGTCTTTTCTTATTA	1267		
Db	51459 ACCATTTAGAGCTTGCAGCAAGTCAGAAACATTTTGTGTGGAAGTCTTTGCTCATGT	51518		
Qy	1268 TGACATGAT-----CGGGGTCAGGTGCTGAAGCTTGAAGGTGGGCTGGGAGTT	1316		
Db	51519 TCATGACGTGCTGATCATCAAGGTGTGTGTGTGTGAAGATGGGGTGGGCTGGGAGTT	51578		
Qy	1317 CT-----TAAACAACAGTGAAGATGCAATATACATTAATCTTCTTCAAGAAAGT	1371		
Db	51579 CTTAAGATATACATACAGTGAAGATTTGCCAATCATTAATCTTCTTCAAGAAAGT	51638		
Qy	1372 TTCTCTGATGTGTGATGCTTTTGTGATAGCATTTTATGCAAGTGAAGCTTTCTTGA	1431		
Db	51639 TTCTCTGATGTGTGATGCTTTTGTGATAGCATTTTATGCAAGTGAAGCTTTCTTGA	51698		
Qy	1432 ATTTGA-TGAATCTCTCAAAACCTGCTGCTTAAACAACCTTAATTAATTAATTC	1490		
Db	51699 ATTTGAGTGAATCTCTTAAAGCTGCTGCTTAAACAACCTTAATTAATTAATTC	51758		
Qy	1491 TGAATCCATTTGTTGCTTCAATTTTCAACATTTTCAACAGTCTTCCAGAGATAGATCCAT	1550		
Db	51759 TGAATCCATTTGTTGCTTCAATTTTCAACATTTTCAACAGTCTTCCAGAGATAGATCCAT	51818		
Qy	1551 CTCATTTCTGAGATGAATCTTTGCTATCCATTAAGAAATTTCTCATCTGTTCAG	1610		
Db	51819 CTCAGAAACCATTTTCTTTGCTTATC-----CAGAAACCATCTCATCTTTCAAG	51873		
Qy	1611 TTTTATCATGAGATTTGAGCAATATACATGCTATGCTTCAAGCTTCACTTTTAATTC	1670		
Db	51874 TTTTATCATGAGATTTGAGCAATATATTTCAATCAACCTTCAAG-----TTCATATTTCTAATTC	51928		
Qy	1671 CAGTCTCTGCTGTTTCAACATCTGTGCTTCTCTGCTGATTTGAAGCTTTGAAGCT	1730		
Db	51929 TGAATCTCTGCTGATTTTCAACATCTGTGCTTCTCTGCTGATTTGAAGCTTTGAAGCT	51981		
Qy	1731 CTCAGATCATCAGAGGTTGGAATCGACTTCTTCAAAATTTCTGTAAATTAATTAAT	1790		

Db	51982 CTCAGATCATCATGAGGTTGGAATCATATTTCTTCCAAACTCTTGTAAATGTTGATAT	52041		
Qy	1791 TTTGACCT---CCCATGAATCATGAATGTTCTTAATGACCTGGAATGGAATCTTT	1847		
Db	52042 TTTGACCTGATGCTTATGATCAAAATATTTCTTAATGATCTGAATGGAATCTTGT	52101		
Qy	1848 CCAAAAGTTTCAATTTACTTAATGTCAGATTCATCCAGAGATTCACCTTTCAATG	1907		
Db	52102 CCAGAAAGTTTTCATCTTACTTGGCCCTGATCAT-----CAGAAATTAAGTGTCTATTC	52156		
Qy	1908 CCAGTTATAGCTTATGGAATGTAATTTCTTCAATATTAAGGCTTGAAGTGAATTAAT	1967		
Db	52157 ATAGGATAGCTTATGGAATGTAATTTCTTCAATATTAAGGCTTGAAGTGAATTAAT	52216		
Qy	1968 CTTGATCCATTTTCTGCAAAATTAAGTGTG 2000			
Db	52217 CTTGGAATGAGGCTTCAAGATGAATGTTGTG 52249			
RESULT 38				
ADN97989_06				
Continuation (7 of 17) of ADN97989 from base 600001 (Human phosphodiesterase 4D genomic				
WP Sequence split into 17 fragments LOCUS ADN97989 Accession Adn97989				
WP	Fragment Name	Begin	End	
WP	ADN97989_00	1	110000	
WP	ADN97989_01	100001	210000	
WP	ADN97989_02	200001	310000	
WP	ADN97989_03	300001	410000	
WP	ADN97989_04	400001	510000	
WP	ADN97989_05	500001	610000	
WP	ADN97989_06	600001	710000	
WP	ADN97989_07	700001	810000	
WP	ADN97989_08	800001	910000	
WP	ADN97989_09	900001	1010000	
WP	ADN97989_10	1000001	1110000	
WP	ADN97989_11	1100001	1210000	
WP	ADN97989_12	1200001	1310000	
WP	ADN97989_13	1300001	1410000	
WP	ADN97989_14	1400001	1510000	
WP	ADN97989_15	1500001	1610000	
WP	ADN97989_16	1600001	1691138	
Query Match 22.1%; Score 441.6; DB 12; Length 110000;				
Best Local Similarity 71.7%; Pred. No. 8.6e-66; Mismatches 232; Indels 66; Gaps 11;				
Matches 755; Conservative 0;				
Qy	973 ATAGTACAGGACATACCTTGAGAGATCTGTGGTTGGTTCCATACACCAATATATACA	1032		
Db	51238 ATATACGAGACATACCTTGAGAGATATTTGGTTCCAGATCCAGGCCACTGCAATTAAGCA	51297		
Qy	1033 AATATGCAAGAGTGAATATCAATTAAGTGCACACAGTCTTTGGCTTCCAGT	1092		
Db	51298 A-----ATATCAAAATTAAGCAACATATATCTTTGGCTTCCAGT	51342		
Qy	1093 GCATATTAAGTTGCTTATCTACCTGCTGCTGTTAAGTGCATATAGTTATGT	1152		
Db	51343 GCATATTAAGTTGCTTATCTACCTGCTGCTGTTAAGTGCATATAGTTATGT	51398		
Qy	1153 CTAAAAAACACATACCTTAATTTTAAATGCTTATTAAT	1207		
Db	51399 AAAACATGATATCTGCTTAAGTAAATATCTTAATTTGCTTAAAAAATGCTTAACA	51458		
Qy	1208 ATCATTTGAGCATGAGTGTGTAATCTTTTGTGTGGAAGTCTTTTCTTATTA	1267		
Db	51459 ACCATTTAGAGCTTGCAGCAAGTCAGAAACATTTTGTGTGGAAGTCTTTGCTCATGT	51518		
Qy	1268 TGACATGAT-----CGGGGTCAGGTGCTGAAGCTTGAAGGTGGGCTGGGAGTT	1316		
Db	51519 TCATGACGTGCTGATCATCAAGGTGTGTGTGTGTGAAGATGGGGTGGGCTGGGAGTT	51578		
Qy	1317 CT-----TAAACAACAGTGAAGATGCAATATACATTAATCTTCTTCAAGAAAGT	1371		
Db	51579 CTTAAGATATACATACAGTGAAGATTTGCCAATCATTAATCTTCTTCAAGAAAGT	51638		

OY	973	ATATGATACAGGCATACCTTGGAGATACCTGAGGAGTTTGCTTGCATATACACACACATATATACA	1032
Db	51238	ATATACCAAGCATACCTCGAGATATTTGGGTTTCAGATTCAGGCCACCTGCAATTAAGCA	5129
OY	1033	AATATGCAAGAAAGTGATATCACATATAAAGTAGTACAACAAGCTCTTTTGGCTTCCAGT	1092
Db	51298	A-----ATATCAAAATTAAGCAAGTACACATACCTTTTGGCTTCCAGT	51342
OY	1093	GCATATAAAGTTTGGTTATCTACACCTAGCTGTGTATAGTGTGCATATAGTGTATGT	1152
Db	51343	GCATATTAACCTTATGTCTACCTATATGTATGTGTGA-----TAGCATTTGCTCTTAAG	51398
OY	1153	CTAAAAAACAACATACCTTATTTTAAATAGCTTTATTACT-----AAAAATGCTTACA	1207
Db	51399	AAAACAATGTACATGCTTATAGTAAAGAAATACTTAAATTTGCTTAAAAAAAATATGCTTACA	51458
OY	1208	ATCATTTGACATTCAGTGAAGTTGTATCTTTTGGTGTGAAGGCTCTTTTCTTATTTGA	1267
Db	51459	ACCATTTAGAGCTTTCAGCAAGTACAGAAACATTTTGGCTGGAGAGATCTTGGCTCAATGT	51518
OY	1268	TGACTGAT-----CGGGGCTCAGGTGCTGAAGCTTAGGGTGGCTGTGGCAGTTT	1316
Db	51519	TCATGACTGCTGACGTATCAAGGTGTGTGTGCTGAAGATTTGGGGGCGCTGTGACATTTT	51578
OY	1317	CT-----TAAACAAACGTGAAGATTTGCAATATACAGTGAAGCTCTTCTTATGAAAGAT	1371
Db	51579	CTTATGATTAACCTACATCAGTAAATTTGCCACATTCATGTATTTCTCTTTCATGAACAT	51638
OY	1372	TTCTCTAGTGTGTGATGCTTTTGTATATGACATTTTATGACAGATGAACCTTCTTGGAA	1431
Db	51539	TTCTCTGTACATGCAATGAGTGTGTGACAGCAATTTTACCATATGATGAACCTTCTTCAAA	51698
OY	1432	ATTGGA-TCATCTCTCAAAACCTGTCTGCTTTAAACAACCTAAGTATATATATATTC	1490
Db	51699	ATTGAGTCAATCTCTTAAACCTGTCTGTGTATATCTAGATTTATATATATTT	51758
OY	1491	TGATTCATGTTGTGATTTCAACATTTTTCACAGTGTCTTACACAGAGATGATTTCCAT	1550
Db	51759	TAAATCTTTTGTGCCATTTCAACAAATGTTACAGACATCTTTTACAGAAAGTATGATCTGT	51818
OY	1551	CTCATTTCTGAGATGGAATCTTTGTGCTCATTCATTAAGAAATAATCTCATCTGTTCAG	1610
Db	51819	CTCAGAAACCATTTTCTTGTTCATTC-----CAAGAAACAACCTCTCATTTCTTCAAG	51873
OY	1611	TTTTATATATGATTTGACGAAATACAGTATGTTCTTCAGGCTCACTTCACTTTATATTC	1670
Db	51874	TTTTATCATTAAGATGACGATTAATTTCTATTACAACCTTCAAG-----TTCCATTTCTAATTC	51928
OY	1671	CAGTTCCTGCGTTTCTTACCAACATCTGTGTGCTCTTCTTCCTCATTTGAAGTGTGAAGCT	1730
Db	51929	TAGTTCTCTTGAATTTCCAGACACATCTGAGTATCTTCTTCA-----CTTGAACCC	51981
OY	1731	CTCCAAGTCATCCATGAGGTTTGAATTCACATCTTCCAAATTCCTGTATATATTAATAT	1790
Db	51982	CTCAGAGTATCCATGAGGTTTGAATTCATTTCTTCCAAACCTCTTGTAAATGTTGATAT	52041
OY	1791	TTTGAAGCT---CCCATGAATCATGAATGTTCTTATGSCACTGTGAATGTGAATCTTT	1847
Db	52042	TTTGAAGCTTAAGCTTAAGATTCACAAATATCTTAAATGAGCATGAGATGTGAATCTGT	52101
OY	1848	CCAAAAGGTTTCAATTTACTTATAGTCCATCATCATCCAGAGAGATCCATTTCAATG	1907
Db	52102	CCAGAAAGTTTGAAGCTTATTTGGCTGATTCAT-----CAAGAAATTAAGTGCTATTC	52156
OY	1908	CCAGTATATAGCTTATGATGATGTATTTCTTCAATATATAAGGCTTGAAGTTGAATTAAT	1967
Db	52157	ATAGCAATATAGCTTATGAAATGTATTTCTTAATATATAGGCTTAAAGTTGAATGACT	52216
OY	1968	CTTGAATCCATTTTCTGCAAAATTAAGTGTGTG	2000
Db	52217	CTTTGAAACATGGGCTTCAAGATGAATGTTGTG	52249

```
RESULT 40
AEB85185_06
Continuation (7 of 17) of AEB85185 from base 600001 (Human phosphodiesterase 4D gene SEQ
WP Sequence split into 17 fragments LOCUS AEB85185 Accession AEB85185
WP Fragment Name Begin End
WP AEB85185_00 1 110000
WP AEB85185_01 100001 210000
WP AEB85185_02 200001 310000
WP AEB85185_03 300001 410000
WP AEB85185_04 400001 510000
WP AEB85185_05 500001 610000
WP AEB85185_06 600001 710000
WP AEB85185_07 700001 810000
WP AEB85185_08 800001 910000
WP AEB85185_09 900001 1010000
WP AEB85185_10 1000001 1110000
WP AEB85185_11 1100001 1210000
WP AEB85185_12 1200001 1310000
WP AEB85185_13 1300001 1410000
WP AEB85185_14 1400001 1510000
WP AEB85185_15 1500001 1610000
WP AEB85185_16 1600001 1691140

Query Match 22.1%; Score 441.6; DB 14; Length 110000;
Best Local Similarity 71.7%; Pred. No. 8.6e-66;
Matches 755; Conservative 0; Mismatches 232; Indels 66; Gaps 11;

QY 973 ATAGTACAGGACCTACCTGGAGATACCTGAGGTTGGTTGCCATACCAACCAATATATACA 1032
DB 51239 ATATACCGAGACCTACCTCGAGAAATATGTTGGTTGCAGATCCAGCCACATCAATTAAGCA 51238
QY 1033 AATATGCAAGAGTGGATATCACAATTAAGTGCACACAAGTCTTTTGGCTTCCAGT 1092
DB 51299 A-----ATATCACAATTAAGCAAGTGCACACAATCTTTTGGCTTCCAGT 51343
QY 1093 GCATATTAAGTTTGGCTTATCTACACTGTAGTCTGTTAAGTGCATATGTTATGT 1152
DB 51344 GCATATTAAGTTTGGCTTATCTACACTGTAGTCTGTTAAGTGCATATGTTATGT 51399
QY 1153 CTAAAAAACAATPCTTAATTTTAAATGCTTATTTACT-----AAAAAAGCTTACA 1207
DB 51400 AAAACAATGTACATGCTTAAAGTAAGAAATCTTAATTTCTTAAAAAATGCTTACA 51459
QY 1208 ATCATTTGAGCATTCAGTGAAGTGTATCTTTTGTGCTGAGGAAGTCTTTTCTTATTTGA 1267
DB 51460 ACCATTTAAGGCTTCAAGCAAGTCAAGAAACATTTTGTGCTGAGGAAGTCTTTGCTCATGT 51519
QY 1268 TGACTGAT-----CGGGGGTCAAGTGTCTGAAGCTTTAGGGTGTGGCTGGCAGTTT 1316
DB 51520 TCATGACTGCTGACTGATCAAGGTGTGCTGCTGAAGATTGGGGTGGCTGTGACATTTT 51579
QY 1317 CT-----TAAACAACAGTAAGATTTGCAATATCACTTACCTTCTTCAATGAAGAT 1371
DB 51580 CTTAAGATTAACACTACAGTAAGAAATTTGCCACATCACTATATCTCTCTTTCATGAACAT 51639
QY 1372 TTCTCTCTAGTGTGATGCTTTTGGATAGCAATTTTATGACAGTAAAGCACTTCTTGAAA 1431
DB 51640 TTCTCTCTAGTGTGATGCTTTTGGATAGCAATTTTATGACAGTAAAGCACTTCTTGAAA 51699
QY 1432 ATTTGA-TCAATCTCTCAAAACCTGCTCTGCTTTAACAACCTTAAGTAAATATATATTC 1490
DB 51700 ATTTGAGTCAATCTCTTAAACGCTGCTGCTGTTATCTACTAATTAATTAATATTT 51759
QY 1491 TCAATTCATTTGTTGATTTTCAACATTTTTCACAGTGTCTTCCACGAGATGATTCAT 1550
DB 51760 TAAATTCCTTTTGGCCATTTCAACATTTTTCACAGATCTTTTACGAAAGTAAATTTCTGT 51819
QY 1551 CTCATTTCTGAGATGAATCTTTGCTCATCCATAGAAGAAATTTCTCATCTGTTCAAG 1610
DB 51820 CTCAGAAACCATTTTCTTTGCTTATC-----CAAGAAACATCTCTCATTTCTTCAAG 51874
QY 1611 TTTTATCATGAGATTGACCAATATACGATGCTTCAAGGCTTCACTTCACTTTAATTC 1670
DB 1611 TTTTATCATGAGATTGACCAATATACGATGCTTCAAGGCTTCACTTCACTTTAATTC 1670
```

```
DB 51875 TTTTATCATGAGATTGACCAATATTTCACTTCAAGS-----TTCCATTTCTAATTC 51929
QY 1671 CAGTTCTCTGCTGTTTCTTACCAATCTGTGTTCTTCTTCCATTAAGTCTGAACTT 1730
DB 51930 TAGTCTCTCTCAATTTCCACACATCTGCAAGTACTTCTTCA-----CTTGAAACC 51982
QY 1731 CTCCAAGTCAATCCATGAGGTTGAATGCACTTCTTCCAAATTCCTGTAAATATATAT 1790
DB 51983 CTCAGAGTCATCCATGAGGTTGAATGCAATTTCTTCCAACTCTTGTAAATGATAT 52042
QY 1791 TTTGACCT---CCCATGATCATGATGTTCTTAATGCACTGGAATGTAATCTTTT 1847
DB 52043 TTTGACCTTATGCTTATGATATCAAAATTTCTTAATGCACTGGAATGTAATCTTGT 52102
QY 1848 CCAAAAGTTTCAATTTACTTACTTACGATTCATTCACAGAGATCCACTTTCAATG 1907
DB 52103 CCAAGAAATTTTCACTTACTTGTGCTGATTCAT-----CAGAGAAATTAAGTGTATTC 52157
QY 1908 CCAATTATAGCTTATGAAATGATATTTTCTTCAATTAATTAAGCTTGAATGAATTTCT 1967
DB 52158 ATAGGATAGCCCTATGAAATGATATTTCTTAATTAATTAAGCTTAAAGTTGAATGACT 52217
QY 1968 CTTGATCCATTTTCTGCAAAATAGATGTTG 2000
DB 52218 CTTGGAACATGGGCTTCAAGATGAATGTTGTG 52250

RESULT 41
ADA02954/C
ID ADA02954 standard; DNA; 96589 BP.
XX
AC ADA02954;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human NR3C1 carcinoma associated gene, SEQ ID NO:1472.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002MO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1472; 245bp; English.
XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
```



CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed human CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;

Query Match 22.1%; Score 441; DB 9; Length 96589;  
Best Local Similarity 71.4%; Pred. No. 1.1e-65;

Matches 733; Conservative 0; Mismatches 230; Indels 64; Gaps 9;

```
QY 974 TAGTACAGGACATCTTGAGATACCTGCGTTGGTTCATACCAACATATATACAA 1033
DB 58294 TAATACAGGACATCTTGAGATATTTTCAGGTTCCGTTTCAGACCAATTA----- 58244
QY 1034 ATATGCAAGAGTGATATCACAAATAAGTACACAAAGTCTTTTGGCTTCCAGTG 1093
DB 58243 -----AAGTAATATGCAACAAAGCAAGTACAGCAATTTTGGTTCCACAT 58193
QY 1094 CATATAAAGTTTGTCTATACCTAGCTAGTCTGTAAAGTGTGCAATAGTATATGTC 1153
DB 58192 CATATAAAGTTTATTTATATACATCTAGCTATCTATAGTGTGCAATAGCTTGTGTC 58133
QY 1154 TAAAAAACAACATACCTTATTTTAAATGCTTATTAATAAATGCTAACATCAT 1213
DB 58132 TTTAAAAATGTAT- -TTACCTTATCTTAAATATTTGCTTAAAAATGCTAACATCAT 58076
QY 1214 TGAGCATTCAGATGTATCTTTTGTGCTGAGAGTCT- - - - -TTTCTTA 1263
DB 58075 TGAATCTTTCAGATGTATCTTTTGTGCTGAGAGTCTTGTGCTGAGTGTATGTC 58016
QY 1264 TTATGATCTGATGAGGAGTCTGAGTCTGAGAGTCTGAGTCTGAGTCTTCTTAAAA 1323
DB 58015 CTGCTGATCTGATGAGGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTTCTTAAAA- 57957
QY 1324 CAACAGTGAAGATGGAATATGAGTCTGAGTCTTCTTCAATGAAAGTCTTCTCAGTGT 1383
DB 57956 -----GTGACAAACATGAAATGCTTCTTCTTCAATGAAAGTCTTCTCAGTGT 57904
QY 1384 TGTATGCTTTTGTATGATGATTTTATGACAGATGAACTTCTTGAATATGAG-ATCAAT 1442
DB 57903 CGAGATGCTGTTGATGATGATTTTACCCAGATGAAATCTTCAAAATGGAATGAAT 57844
QY 1443 CCTCTGAAACCTGCTGCTGCTTAAACAGCTTAAGTAAATATTTGATGATCAATGT 1502
DB 57843 CCACCTGAAACATGCTGCTGCTTAAACAGCTTAAGTAAATATTTGATGATCAATGT 57784
QY 1503 TGTATTTTGAACAAATTTTCAAGTCTTCAACAGAGATGATGATTCATCTCATTTCTCA 1562
DB 57783 TGTATTTTGAACAAAGTTTCAAGATCTTCAACAGAGATGATGATTCATCTCAAGAAAAA 57724
QY 1563 GATGATCTTTTGTCTATCATTAAGAAATTTCTCATGTTTCAAGTTTATTCATGAG 1622
DB 57723 ATTTATTTGCTCATTTTGTCTATTAAGAAATCTCTCATGATTCATGAGTTTCTCATGAG 57664
QY 1623 ATTTGCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1682
DB 57663 ATTTGCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 57609
QY 1683 TGTATTTTGAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1742
DB 57608 TGTATTTTGAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 57549
QY 1743 CATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1801
DB 57548 TATATGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 57489
```

```
QY 1802 ATGAATCAGTAATGTTCTTATGACCTGAGATGATGATGATGATGATGATGATGAT 1861
DB 57488 CTGCGAATACAAATTTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 57429
QY 1862 ATTTACTTATGTCAGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1921
DB 57428 ATTTACTTATGTCAGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 57374
QY 1922 ATGGAATGATTTTCTTCAATTAATGAGCTTGAAGTGAATATTTCTTCAATTCATTT 1981
DB 57373 ACMAATGATCT- - - - -GATTTGAAACTTGAATACATCTCTGATTCATGAG 57327
QY 1982 CTGCAAA 1988
DB 57326 CTACAAA 57320
```

RESULT 42

ADB72692/c  
ID ADB72692 standard; DNA; 96589 BP.

XX ADB72692;

DT 04-DEC-2003 (first entry)

XX Human NR3C1 gene.

XX human; ds; cytosolic; gene therapy; vaccine; carcinoma; lymphomas;

XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-0004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX MPI; 2003-239337/23.

XX Claim 1; SEQ ID NO 520; 2304bp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a

XX nucleotide sequence selected from any of the 660 sequences fully defined

XX in the specification. A polynucleotide of the invention has cytosolic

XX activity, and may have a use in gene therapy, or in a vaccine. The

XX recombinant nucleic acids and polypeptides are useful for treating

XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and

XX sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;

Query Match 22.1%; Score 441; DB 10; Length 96589;

Best Local Similarity 71.4%; Pred. No. 1.1e-65;

Matches 733; Conservative 0; Mismatches 230; Indels 64; Gaps 9;

```
QY 974 TAGTACAGGACATCTTGAGATACCTGCGTTGGTTCATACCAACATATATACAA 1033
DB 58294 TAATACAGGACATCTTGAGATATTTTCAGGTTCCGTTTCAGACCAATTA----- 58244
```

```

QY 1034 ATATGCAAGAGTGATATCAATATTAAGTGACACAAAGTCCTTTGGCTTCCAGTG 1093
Db 58243 -----AGTGAATATGCAACAAAGCAAGTCAGACGAATTTTGGTTCCACATA 58193
QY 1094 CATATAAAAAGTTTGGTTATCTATCACTGTAAGTCTGTTAAGTGTGCAATAGTATATGTC 1153
Db 58192 CATATAAAAAGTTATTTATCTATCACTGTAAGTCTGTTAAGTGTGCAATAGTATATGTC 58133
QY 1154 TAAAAAACAATACCTTAATTTTAAATGCTTTTATTAATAAATGCTAACATCAAT 1213
Db 58132 TTTAAATAATGTAT---TTACCTTAATCTTAATAATTTGCTTTAAATGCTTAACAATCAT 58076
QY 1214 TGAGCATTCAGTAGTGTGTAATCTTTTGTGCTGTGGAGAGTCT-----TTTCTTA 1263
Db 58075 TGAATCTTCAGTAGTGTGTAATCTTTTGTGCTGTGGAGAGTCTGTGCTTCAGTGTGATGC 58016
QY 1264 TTGATGACTGATCGGGGGTCAAGTGTGTAAGCTTAAAGGCTGTGGCAGTTCTTAA 1323
Db 58015 CTGTGACTGATCAGAGAGGTGGTGTCTGAAGGTGGGGTGGCTGTGGCAATTTCTTAA 57957
QY 1324 CAACAGTAGATGCAATATAGTTGACTCTTCCCTTATGAAGATTTCTCTAGTG 1383
Db 57956 -----GTAGCAACAATGAAGTTTCACTTTTCTTTATAAAGATTTCTTTGAGCA 57904
QY 1384 TGTGATGCTTTTGTATAGCATTTTATGACAGTAGAATCTTTGAAAATTTGG-ATCAAT 1442
Db 57903 CGAATGCTGTTGTATAGCATTTTACCCACAGTAGAATCTTTGAAAATTTGGAATGAT 57844
QY 1443 CCTTCAAAACCTGCTGCTGCTTTAACACCTTAAATTAATTAATTTGAAATCCATTGT 1502
Db 57843 CCACTCAAACATGCTGCTGCTTTTACCAATTAATTAATTAATTTGAAATCCATTGT 57784
QY 1503 TGTCAATTTCAACAATTTTACAGTGTCTTCAACAGAGTAGATTCATCTCATTTCCCTGA 1562
Db 57783 TGTATTTTCAACAAGTTTACAGATTTTCAACAGAGTAGATTTCAATCTCAAGAAAAA 57724
QY 1563 GATGAAATCTTGTCTATCATTAAGAGAAATTTCTCATCTGTCTCAAGTTTATCATGAG 1622
Db 57723 ATTTATTTGCTCATTTGCTCATTAAGAGCACTCCATCAATGACGTTTCTCATGAG 57664
QY 1623 ATTTGACAAATACAGTCATGCTCTTCAAGGCTTCACTTCAATTTTCAAGTCTTCTGC 1682
Db 57663 ATTTGACAAATTCCTGCAATTTTCCGG-----CTCCACTTCTAATTTCTAGTCTTCTGC 57609
QY 1683 TGTTCCTACACATCTGTGCTCTTCCCTCCATGAAGTCTGAACCTCTCAAGTCATC 1742
Db 57608 TGTTCCTACACATCTGTGCTCTTCCCTCCATGAAGTCTGAACCTCTCAATTCATC 57549
QY 1743 CATGAGGTTTGAATGACCTTCTTCAAAATTCCTGTTAATTTATTTATTTTGAACCTCC-C 1801
Db 57548 TATTAATCTTGGATTAATCTTCTTCAAACTGTACTAATGTTGATCTTTGACCTCCGC 57489
QY 1802 ATGAATCATGAATGTTCTTAATGACCTGTGAATGCTGAATGCTTTTCAAAAAGTTTCA 1861
Db 57488 CTGCGAATCAACAATCTTAATGACATCTGAATGCTGAATGCTTTTCAAAAAGTTTTCG 57429
QY 1862 ATTTCTTGTGCAAGTTCATCATGAGAGATCCATTTTCAATGCCAGTTATATACCTT 1921
Db 57428 ATTTCTTGTGCAAGTTCATCATGAGAGATCCATGCTTAATGAGCTTAATACCTT 57374
QY 1922 ATGAATGATTTTCTTCAATTAATGAAGCTTGAAGTGAATTAATCTTGAATCTCTTTC 1981
Db 57373 ACAAATATGATCT-----GATTTGAAATCTTGAATATCACTCCCTGATTCATGG 57327
QY 1982 CTGCAAA 1988
Db 57326 CTACAAA 57320

```

RESULT 43  
 ADC85434/C  
 ID ADC85434 standard; DNA; 96589 BP.

```

XX AC ADC85434;
XX DT 01-JAN-2004 (first entry)
XX DE Human NR3c1 genomic sequence.
XX KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
XX KW secreted; transmembrane; intracellular; ds.
XX OS Homo sapiens.
XX EN WO2003045230-A2.
XX PD 05-JUN-2003.
XX PF 02-DEC-2002; 2002MO-US038582.
XX PR 30-NOV-2001; 2001US-00997722.
XX PA (SAGR-) SAGRES DISCOVERY.
XX PI Morris DW, Engelhard EK;
XX DR WPI, 2003-513603/48.
XX PT New recombinant nucleic acid comprising a nucleotide sequence of any of
XX PT the carcinoma-associated (CA) genes, useful for screening for drug
XX PT candidates for diagnosing or treating carcinomas.
XX PS Claim 1; SEQ ID NO 220; 983bp; English.
XX CC The invention relates to a recombinant nucleic acid comprising a
XX CC nucleotide sequence selected from any of the fully defined carcinoma-
XX CC associated (CA) genes from the 50 tables given in the specification. The
XX CC CA proteins are secreted, transmembrane or intracellular proteins. The
XX CC recombinant nucleic acids are useful for screening for drug candidates
XX CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
XX CC ADC85514 represent CA genes of the invention.
SQ Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;
Query Match 22.1%; Score 441; DB 10; Length 96589;
Best Local Similarity 71.4%; Pred. No. 1.1e-65;
Matches 733; Conservative 0; Mismatches 230; Indels 64; Gaps 9;
QY 974 TAGTACAGGCAATCTTGGAGATGATCTGTGGTGGTTCCATACCAACAATAATACAA 1033
Db 58294 TAAATCAGGCAATCTTGGAGATGATGATCTGTGGTGGTTCCATACCAATA----- 58244
QY 1034 ATATGCAAGAGTGATATCAATATTAAGTGACACAAAGTCCTTTGGCTTCCAGTG 1093
Db 58243 -----AGTGAATATGCAACAAAGCAAGTCAGACGAATTTTGGTTCCACATA 58193
QY 1094 CATATAAAAAGTTTGGTTATCTATCACTGTAAGTCTGTTAAGTGTGCAATAGTATATGTC 1153
Db 58192 CATATAAAAAGTTATTTATCTATCACTGTAAGTCTGTTAAGTGTGCAATAGTATATGTC 58133
QY 1154 TAAAAAACAATACCTTAATTTTAAATGCTTTTATTAATAAATGCTAACATCAAT 1213
Db 58132 TTTAAATAATGTAT---TTACCTTAATCTTAATAATTTGCTTTAAATGCTTAACAATCAT 58076
QY 1214 TGAGCATTCAGTAGTGTGTAATCTTTTGTGCTGTGGAGAGTCT-----TTTCTTA 1263
Db 58075 TGAATCTTCAGTAGTGTGTAATCTTTTGTGCTGTGGAGAGTCTGTGCTTCAGTGTGATGC 58016
QY 1264 TTGATGACTGATCGGGGGTCAAGTGTGTAAGCTTAAAGGCTGTGGCAGTTCTTAA 1323
Db 58015 CTGTGACTGATCAGAGAGGTGGTGTCTGAAGGTGGGGTGGCTGTGGCAATTTCTTAA 57957
QY 1324 CAACAGTAGATGCAATATAGTTGACTCTTCCCTTATGAAGATTTCTCTAGTG 1383
Db 57956 -----GTAGCAACAATGAAGTTTCACTTTTCTTTATAAAGATTTCTTTGAGCA 57904

```

QY 1384 TGTGATGCTTTTGTAGATTTTATGACAGTAGAATCTTTGAAATTTG-ATCAAT 1442  
DB 57903 CAGAGATGCTTTGTATAGATTTTATCCACAGTAGAATCTTTCAAACTGAAAT 57844  
QY 1443 CCTTCAAACTGCTCTGCTTTAAACACCTAAATTAATTAATTTGAAATTCAT 1502  
DB 57843 CCAGTCAAACTGCTGCTGCTTTACCAATTAATTAATTTGAAATTCAT 57784  
QY 1503 TGTGATTTGAAATTTTCAAGTCTTTCAAGAGATGATTCATCTCATTTCTGA 1562  
DB 57783 TGTATTTTCAACAAAGGTTCACAAATCTTCACAGAGATGATTCATCTCA 57724  
QY 1563 GATGATCTTTTCTGCTCATTAAGAAATCTTCATCTGTTCAAGTTTATCAAG 1622  
DB 57723 ATTTATTTGCTCATTTGCTATTAAGAAAGCACTTCATTCATTCAGTTTCT 57664  
QY 1623 ATTGAGCAATACAGTCAATGCTTCAAGGCTCACTTCATTTAATTCAGTT 1682  
DB 57663 ATTGAGCAATTCGTCATCTTTGAGG-----CTCAGTCTAATTTAGTTCT 57609  
QY 1683 TGTGTTTACCAATCTGCTGCTTCTTCATTTGAAGTCTTGAACCTTCAAG 1742  
DB 57608 TGTGTTTACCAATCTGCAAGTCACTTCATTAAGTGTGAACCTTCAAAAT 57549  
QY 1743 CATGAGGTTTGAATGAGCTTCTTCAAAATCTGTTAATTTATTTGACCTCC 1801  
DB 57548 TATTAATCTTTGAAATCACTTCTTTCAAACTGCTAATTTGATCTTTGAC 57489  
QY 1802 ATGAAATCATGAATGTTCTTAATGAGCACTGGAATGAGTGAATCTTCA 1861  
DB 57488 CTGCAATCAAAATCTTTAATGAGCAATGAGCAATCTTTTGAAGTTTGG 57429  
QY 1862 ATTTACTTATGCTCAATTCATTCATTCAGAGATTCATCTTCAATGCTT 1921  
DB 57428 ATTTACTTATGCTCAATTCATTCAGAGATTCATCTTCAATGCTT 57374  
QY 1922 ATGGAATGATTTCTTCAATTAAGCTTGAAGTGAATTTACTCTGATCC 1981  
DB 57373 ACAAATGATCT-----GATTTGAAATTTGAATCACTCCGATTCATG 57327  
QY 1982 CTGCAAA 1988  
DB 57326 CTACAAA 57320

RESULT 44  
ADM74549/c  
ID ADM74549 standard; DNA; 96589 BP.  
XX  
AC ADM74549;  
XX  
DT 01-JUL-2004 (first entry)  
XX  
DE Human carcinoma associated (CA) nucleic acid #109.  
XX  
KW Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds,  
XX  
KW carcinoma associated protein; CAP; carcinoma; leukemia; lymphoma;  
XX  
OS Homo sapiens.  
XX  
XX US2004072154-A1.  
XX  
PN 15-APR-2004.  
XX  
PD 30-NOV-2001; 2001US-00997722.  
XX  
PF 22-DEC-2000; 2000US-00747377.  
XX  
PR 02-MAR-2001; 2001US-00798586.  
XX  
XX (MORRIS D W.  
XX  
XX (ENGELHARD E K.

XX  
PI Morris DW, Engelhard EK;  
XX  
DR WPI, 2004-328562/30.  
XX  
PT New carcinoma associated gene or protein, useful for preparing a  
PT composition for diagnosing or treating carcinoma e.g., leukemia or  
XX lymphoma.  
PS Claim 1, SEQ ID NO 220; 299p; English.  
XX  
XX The invention relates to new recombinant nucleic acids. The invention  
CC also relates to a host cell comprising a recombinant nucleic acid or  
CC expression vector, an expression vector comprising a recombinant nucleic  
CC acid, a recombinant protein, a method of screening for drug candidates, a  
CC method of screening for a bioactive agent capable of binding to a  
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a  
CC method of screening for a bioactive agent capable of modulating the  
CC activity of a CAP, a method of evaluating the effect of a candidate  
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting  
CC the activity of a CAP, a method of treating carcinoma, a method of  
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or  
CC propensity to carcinoma. A method of evaluating the effect of a candidate  
CC carcinoma drug comprises administering the drug to a patient, removing a  
CC cell sample from the patient and determining alterations in the  
CC expression or activation of a gene comprising the nucleotide sequence. A  
CC method of diagnosing carcinoma comprises determining the expression of  
CC one or more genes comprising the nucleic acid sequence in a first tissue  
CC type of a first individual and comparing the expression of the gene from  
CC a second normal tissue type from the first individual or a second  
CC unaffected individual, where a difference in the expression indicates  
CC that the first individual has carcinoma. A method of inhibiting the  
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating  
CC carcinoma comprises administering to a patient an inhibitor of CAP.  
CC Neutralising the effect of a CAP comprises contacting an agent specific  
CC for the CAP. The polypeptide specifically binds to the protein encoded by  
CC the nucleic acid. It comprises an antibody that specifically binds to the  
CC protein encoded by the nucleic acid. The nucleic acids are useful for  
CC preparing a composition for diagnosing or treating carcinoma e.g.,  
CC leukemia or lymphoma. This sequence represents a human carcinoma  
CC associated (CA) nucleic acid of the invention. Note: The sequence data  
CC for this patent did not form part of the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 96589 BP; 28270 A; 17369 C; 18773 G; 31440 T; 0 U; 737 Other;  
XX  
Query Match 22.1%; Score 441; DB 12; Length 96589;  
Best Local Similarity 71.4%; Pred. No. 1.1e-65;  
Matches 733; Conservative 0; Mismatches 230; Indels 64; Gaps 9;  
QY 974 TAGTACAGGATACCTTGAGATGAGTGTGGTTGCTTACACCAATTAATCA 1033  
DB 58294 TAATACAGGATACCTTGAGATGAGTGTGGTTGCTTACACCAATTA 58244  
QY 1034 ATATGCAAGATGATATCAATTAAGTGTGTACACAGTCTTTGGCTTCCAGTG 1093  
DB 58243 -----AGTGAATATTTGCAACAAAGCAAGTCAAGATTTTGGTTCCACAT 58193  
QY 1094 CATATTAAGTTTGTATATACATACAGTGTAGTCTGTTAAGTGAATAGTTATGTC 1153  
DB 58192 CATATTAAGTTTGTATATATATATATATATATATATATATATATATATAT 58133  
QY 1154 TAAATAAACAATACCTTAATTTTAAATGCTTTAATTAATAAATGTAACAATCAT 1213  
DB 58132 TTTAATAAATGATATTTTACCTTAATTTTAAATGCTTTAATAAATGTAACAATCAT 58076  
QY 1214 TGAGATTCAGAGTGTGAATCTTTTGTGCTGAGAGTCT-----TTTCTTA 1263  
DB 58075 TGAATCTTCAGAGTGTGAATCTTTTGTGCTGAGAGTCTTTTCTTCAAGTGTGAATGC 58016  
QY 1264 TTGATAGCTGATCGGGGGGTCAAGTGTGAAGCTTGAAGGTTGCTGTGAGAGTTCTTA 1323

Db	58015	CTGCTGACTGATTCAGAGAGAGGTTGCTGTGAAGGTGGGGTGGCTGTGGCAATTTCTTAAT-	57957
Qy	1324	CAACAGTGAAGATTGCAATATTCAGTTGACTCTTCTTTCATGAAAGATTTCTCTTAGTG	1383
Db	57956	-----GTAGACAACACATGAAAGTTTCACTTTCTTTTCAATAAAGATTTCTTTGAGCA	57904
Qy	1384	TGTCGATGCTTTTGTATATGCAATTTATGACAGATGAACTTCCTTTGAAATTTGG-ATCAT	1442
Db	57903	CGAATGCTGTTTATATGACATTTTACCCACAGTGAAGAACTCTTTCAAAACCTGAATGAAT	57844
Qy	1443	CCTCTCAAAACCCCTGCTCTGCTTTTAAACAACCTAAGTTAATATTAATTTCTGAATCCATTGT	1502
Db	57843	CCACTCAAAACCATCTGCTGCTGTTTACCAATTAAGTTATATGTAATCTTAATCCTTTGT	57784
Qy	1503	TGTCATTTTCAAACAATTTTCAACAGTCTTTCACACAGAGATGATTCATCTCAATTTCTGA	1562
Db	57783	TGTTATTTTCAACAAGTTTTCACAGAACTTTCACACGAGAGTAGATTTCATCTCAAGAAAAA	57724
Qy	1563	GATGGAATCTTTGGTCACTCCATTAAGAGAAATTCCTATCTGTCAAGTTTATTCATGAG	1622
Db	57723	ATTATATTTGCTCATTTTGTCCATTAAGAGAACTTCATCATTCACGTTTTCTCAVAG	57664
Qy	1623	ATTGACAGATATACAGTCAATGCTTCAGAGCCATCACTTCACTTTAATTTCCAGTCTCTTGC	1682
Db	57663	ATTGACAGAAATCCGTCACATTTTCGG-----CTCACCTTTAATTTCTAGTTCTCTTGC	57608
Qy	1683	TGTTTTCTACACATCTGTGTGTTCCCTTCCTCCATTGAAGTCTTAACCTTCCAAATCATC	1742
Db	57608	TGTTTCTACACATCTGTGAGTGACTTCTCTAACTGAAGTGTGAACCCCTCAATTCATC	57548
Qy	1743	CATGAGGTTTGGAATCCACTTCTTCCAAATTCCTGTTAATATTTAATTTTGAATCC--C	1801
Db	57548	TATATCTGTTGAAATCAACTTCTTTCAAACGTCACTAATGTTGATCTTTGACCTTCCTC	57489
Qy	1802	ATGATATCATGAATGTTCTTAATGGCACCTGGAATGTGAATTCCTTCCAAAAGTTTTC	1861
Db	57488	CTGCGAATCACAAAATTTCTTAATAGACTCTAGAAATGGCAATCTTTTCAGAGATTTTG	57422
Qy	1862	ATTTACTTATGTCAGATCCATCCATCCATCCAGAGATCCACTTTCATATGCACTTAATAGCTTT	1921
Db	57428	ATTTATCTTTGGCCAGGTCCA-----CCAGGACATCACTGCTATAGCAGCTAATAGCTTT	57374
Qy	1922	ATGGAATGTATTTCTTCAATATATAGGCTTGAAAGTTGAATTAATCTCTTGATCCATTTT	1981
Db	57373	ACAAATGTACT-----GATTTGAAACTTGAATAATCACTCCGATTCATAGG	57322
Qy	1982	CTGCAAA 1988	
Db	57326	CTACAAA 57320	
RESULT 45			
ADFL13116			
ID	ADFL13116 standard; DNA; 128779 BP.		
AC	ADFL13116;		
XX			
DT	12-FEB-2004 (first entry)		
XX			
DE	Hypermethylation site in human breast cancer CpG island locus HBC-45.		
XX			
KW	DNA methylation; CpG dinucleotide rich region; CpG island; screening array;		
KW	differential methylation hybridisation; DMH; CpG island; screening array;		
KW	breast cancer; prostate cancer; colon cancer; lung cancer; liver cancer;		
XX	ovarian cancer; human; HBC; hypermethylation in breast cancer; ds.		
OS	Homo sapiens.		
XX			
PN	US2003129602-A1.		
XX			
PD	10-JUL-2003.		
XX			
PF	21-FEB-2002; 2002US-00081327.		

XX	05-FEB-1999;	99US-0118760P.	
PR	18-FEB-1999;	99US-0120582P.	
PR	04-FEB-2000;	2000US-00497855.	
XX			
PA	(HUAN/)	HUANG T H.	
XX			
PI	Huang TH;		
XX			
DR	WPI; 2004-031298/03.		
XX			
PT	Differential methylation hybridization assay for detecting		
PT	hypermethylation of CpG dinucleotide rich regions in genomic DNA is		
PT	useful to diagnose and monitor breast, prostate, colon, lung, liver and		
PT	ovarian cancer.		
XX			
PS	Claim 17; SEQ ID NO 38; 32pp; English.		
XX			
CC	The present invention relates to a method of detecting DNA methylation of		
CC	a CpG dinucleotide rich region of a nucleic acid. The method comprises		
CC	hybridization by differential methylation hybridisation (DMH) in which		
CC	the nucleic acid is digested into fragments in which CpG islands are		
CC	preserved, attaching the fragments to linker primers, digesting the		
CC	linker primer products to digest unemethylated CpG sequences, amplifying		
CC	and labelling the remaining linker primer products and detecting labelled		
CC	amplicons. Also disclosed is a screening array comprising nucleic acid		
CC	fragments fixed to a solid support where each fragment is a CpG		
CC	dinucleotide rich fragment comprising at least 200 nucleotides of which		
CC	at least 50% are guanine and cytosine. Nucleic acid probes prepared from		
CC	a cell sample are used to screen CpG dinucleotide rich fragments fixed		
CC	onto a screening array. The invention is useful to diagnose and monitor		
CC	prognosis of a disease associated with aberrant DNA methylation.		
CC	particularly breast, prostate, colon, lung, liver or ovarian cancer,		
CC	especially breast cancer. The present sequence represents the site of		
CC	hypermethylation in human breast cancer at a CpG island locus (HBC,		
CC	hypermethylation in breast cancer).		
XX			
XX	Sequence 128779 BP; 38014 A; 25995 C; 25782 G; 38988 T; 0 U; 0 Other;		
Query Match	22.0%; Score 439.8; DB 12; Length 128779;		
Best Local Similarity	72.7%; Pred. No. 1.7e-55;		
Matches 747; Conservative	0; Mismatches 232; Indels 48; Gaps 12;		
QY	977 TACAGCATACCTTGGAGATACCTGTGGGTTTGCTTCATACCAACCATATATACAAATTA	1036	
DB	42749 TACAGGCATACATTTGGACATATTTGGCTTTAGACCACTGCACAAAAGCAAAATTA	42808	
QY	1037 TGCAAGAAAGTGATATCAACATTAAGAGAGTGCACACAAGCTTTGGCTTCCAGAGCAT	1096	
DB	42809 TTG-----CAATTAAGTGAATTCACAAATTTTTTTCCTTTCCAGTGCAT	42853	
QY	1097 ATAAAGTTTGCCTTATATACACTGTAGTCTGTTAAGTGTGCATATAGTTATGCTTAA	1156	
DB	42854 ATAAAGTTATGTTTATTAACATACAGCATATTAAGTGTGCATAGCAT--GTTCTAA	42911	
QY	1157 AAAAAACATACCTTAATTTTAAATGCTTTATTTACTTAAATAATGCTAACATCATTTGA	1216	
DB	42912 AAATGCAATATCCTTAATTAATAAGATCTTATTTGCTTAAATAATG-CAGACTCTTCACT	42970	
QY	1217 GCATTCAGTAGTGTATCTTTTGTGTGGAAGGCTCTTTCTTATATGATGATGCA-T	1275	
DB	42971 GAGTATATCTTTTGAAGCTGCAGAGCTTAACCTTAATGCTGTCTGACTCATAGGCTGAT	43030	
QY	1276 CGGGGCTCAGGCTGCGAAGCTTAAAGGCTGCTGCGAGCTTTCTTAA-----ACAACT	1330	
DB	43031 AGCGTGATTAATCTGCGAAGGCTCAGGCTGCGAGCTTTCTTAAACATGACAACT	43090	
QY	1331 GAAGATTGCAATATCACTGCTCTTCTTTCAATGAAGATTTCTCTTAGTGTGTGATG	1390	
DB	43091 GAAGTTTCTGCTCATCAGTTGATCTTCTCTTCAATGAACATATTTCTCTGTGTATGTGAAG	43150	
QY	1391 CTTTGTATACATTTTATGACAGTGAACCTTCTTTGAAAATTTGATCAATCTCTGAA	1450	

```
Db 43151 CTGTTAATAGCATTTTACCCAGAA-----CTCTAAAGTTGAGTCATCTCTCAA 43205
QY 1451 ACCCTGCTGCTTTTAAACAACCTAAGTTATATATATTTGTGAATCCATTGTTGCAATT 1510
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43206 ACTTGCCCACTGGTTTATCAAGTCAGATTTATGTGACATTCCTAATCTTGGCTATATT 43265
QY 1511 CAACAATTTTCAAGAGTCTTCAACGAGATGATTCATCTCATTTCTCTGAGATGGAAT 1570
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43266 CAACAA-TTTCACAGCATCTTCCCGAGAGTAGATTTCATCCFAGAAACCACTT----T 43320
QY 1571 CTTTGTCTCATTAAGANGAAMATTCCTCATCTGTTCAAGTTTATCATGAGATTGCACG 1630
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43321 CTTTACTCATTCATAGAGCAACTTCATCCATTCAAATTTTATCATAGATGCAGC 43380
QY 1631 AATACAGTATGCTTTCAGGCCCTCACTTCACTTTTATTCAGTCTCTTGGCTTTCTA 1690
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43381 AATTCGTACACCTTCAGG-----CTACACTTCTAATCTAGTCTCTGCTATTTCCA 43435
QY 1691 CCACATCTGAGTCTCTCTCCATGTAAGTCTTGAACCTTCCAAGTCATCCATGAGGG 1750
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43436 CTRATCTTACAGTTATTTCTCTCAGTGAAGTCTTGAACCTGGAAGTCATCCATGAGGA 43495
QY 1751 TTGGAATGACTTCTTCCAAATTCCTGTTAATAATTTTGA---CCTCCCATGAAT 1807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43496 TGGGAATCAACTTCTTCCAACTCCTGTTCAIGTGAATTTTGAATCTCTCCCATGAAT 43555
QY 1808 CATGAATGTTCTTAATGCACTGGAATGATGATCTTCCAAAAGTTTCAATTTAC 1867
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43556 CACAAATGTTCTTAATGATGATCTAGAAAT-GTGAATCTCTCCAGAAAGTTTCAATGTAC 43614
QY 1868 TTAGTCAGATCCATCCATCCAGAGATCCACTTTCATGCCAGTATAGCCTTAAGAA 1927
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43615 TTTGCTCAGATCCAT-----CAGAGGAATCACTATCTATGCAAGCTATACCTTACAAA 43669
QY 1928 TGTATTTCTTCAATATAAGGCTTGAAGTTGAATTTACTCTGTGATCCATTTTCGCA 1987
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43670 TGTATTTCTTAATATAAGACTTGAAGTTGAATTTACTCTGTGATCTGTGGGCTGTGG 43729
QY 1988 AATAGAT 1994
    |||||
    |||||
Db 43730 AATGGTT 43736
```

Search completed: January 22, 2006, 04:08:38  
Job time : 831.678 secs

This Page Blank (uspto)







Db 1681 GCTGTTTACACACATCTGCTGCTTCCCTCCATGGAAGCTTGAACCTCCCAAGTCA 1740  
Qy 1741 TCCATGAGGGTGGAAATGCACTCTCCAAATTCCTGTAATATTAATTTGACCTCC 1800  
Db 1741 TCCATGAGGGTGGAAATGCACTCTCCAAATTCCTGTAATATTAATTTGACCTCC 1800  
Qy 1801 CATGATCATGATGATGTTCTTAATGAGCACTGGAATGGAATCTTCCAAAAGTTTTC 1860  
Db 1801 CATGATCATGATGATGTTCTTAATGAGCACTGGAATGGAATCTTCCAAAAGTTTTC 1860  
Qy 1861 AATTACTAGTCAGATCCATCCATCCAGAGATCCATCTTCAATGCCATTAAGCCT 1920  
Db 1861 AATTACTAGTCAGATCCATCCATCCAGAGATCCATCTTCAATGCCATTAAGCCT 1920  
Qy 1921 TATGAAATGATTTCTTCAATAATAAGGCTGGAAGTTGAATTAATCTCTGATCCATTT 1980  
Db 1921 TATGAAATGATTTCTTCAATAATAAGGCTGGAAGTTGAATTAATCTCTGATCCATTT 1980  
Qy 1981 TCTGCAAAATAGATGTTGTG 2000  
Db 1981 TCTGCAAAATAGATGTTGTG 2000

RESULT 2  
US-09-078-294-4  
; Sequence 4, Application US/09078294  
; Patent No. 6255211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 80246  
; TYPE: DNA  
; ORGANISM: Nucleotide sequence of NC-contig  
US-09-078-294-4

Query Match 80.6%; Score 1611.4; DB 3; Length 80246;  
Best Local Similarity 97.4%; Pred. No. 0;  
Matches 1666; Conservative 2; Mismatches 40; Indels 3; Gaps 3;

Qy 292 GATTGATCTGCTATGATGATGCTGATTCATTCCTCGATAGTATTTCTTTTCTAAT 351  
Db 2 GGTGATTTGTNNATAGAGAGTTTGATCAATCCCGAGAGAAATTTTITTTTAAAA 61  
Qy 352 ATTTTTCGAATGCTCTGCTATTAACATAGCCACTGCGCTTTTAAAAATAGTATTT 411  
Db 62 ATTTTTCGAAGG-GTTTGTATTAATAAARCAATTTGGCTTTTAAAAATAGAAATTT 120  
Qy 412 TATGTAATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 471  
Db 121 TATGGAATAAAATTTTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 180  
Qy 472 TTAAGTGGGTGCTTATAGGAGCATATATCGGGCTTGTAGTATTAATTAATCTGAT 531  
Db 181 TTAAGTGGGTGCTTATAGGAGCATATATCGGGCTTGTAGTATTAATTAATCTGAT 240  
Qy 532 AATCTCAACCTTTTGTGGAAGTGTTAAGGCAATTTAGTGTAAATTAAGACATG 591  
Db 241 AATCTCAACCTTTTGTGGAAGTGTTAAGGCAATTTAGTGTAAATTAAGACATG 300  
Qy 592 GTTTGATTTGCTAATCCATCTTTCACTTTTATATAGGAGCATCTTTTCAATGTC 651  
Db 301 GTTTGATTTGCTAATCCATCTTTCACTTTTATATAGGAGCATCTTTTCAATGTC 360  
Qy 652 TTTTTCATCTTTGACCATTTTCTTTAGTGAATACTTTTCTTTTCTTTTCTTTTCT 711

Db 361 TTTTTCATCTTTGACCATTTTCTTTAGTGAATACTTTTCTTTTCTTTTCTTTTCT 420  
Qy 712 TATGAGCTTTTATGATATACCTTAATTTTTTTTTTCTGTTTATATAGATTTATATA 771  
Db 421 TATGAGCTTTTATGATATACCTTAATTTTTTTTTTCTGTTTATATAGATTTATATA 480  
Qy 772 TATGATCTTTTATCTTATACAGATATTAACCTTAATATATATTTTACAGCTCAAGTATA 831  
Db 481 TATGATCTTTTATCTTATACAGATATTAACCTTAATATATATTTTACAGCTCAAGTATA 540  
Qy 832 TGTAAACCTTAACAAGATATATTTTCAATTTCTGCTCTAATTTTATAGCTA-TGTCT 890  
Db 541 TGTAAACCTTAACAAGATATATTTTCAATTTCTGCTCTAATTTTATAGCTATGCTG 600  
Qy 891 ATATATCATATAGTTTGTGTTGTTTGTGTTTATACCTATATGCTGTGGCTGGGTACGA 950  
Db 601 ATATATCATATAGTTTGTGTTGTTTGTGTTTATACCTATATGCTGTGGCTGGGTACGA 660  
Qy 951 AACATTTTCTGTAAGGGCTAGATAGTACAGCATATCTTGGAGATACGTGGGTTTGGT 1010  
Db 661 AACATTTTCTGTAAGGGCTAGATAGTACAGCATATCTTGGAGATACGTGGGTTTGGT 720  
Qy 1011 TCCATACCAACAATAATATACAAATATGCAAGAAGTATATCACAAATAGTGAATCAC 1070  
Db 721 TCCATACCAACAATAATATACAAATATGCAAGAAGTATATCACAAATAGTGAATCAC 780  
Qy 1071 ACAAGCTTTTGGCTTCCAGTGCATATTAAGTTTGTCTTATATCACTGATGCTGT 1130  
Db 781 ACAAGCTTTTGGCTTCCAGTGCATATTAAGTTTGTCTTATATCACTGATGCTGT 840  
Qy 1131 TAAAGTGCATATAGTTATATGCTAATAAACAACATCTTAATTTTAAATAGCTTAT 1190  
Db 841 TAAAGTGCATATAGTTATATGCTAATAAACAACATCTTAATTTTAAATAGCTTAT 900  
Qy 1191 ACTAAAAATGCTAAACAATCATTTAGAGCATTCAGTGAAGTGTATCTTTTCTGGTGA 1250  
Db 901 ACTAAAAATGCTAAACAATCATTTAGAGCATTCAGTGAAGTGTATCTTTTCTGGTGA 960  
Qy 1251 AGGCTTTTCTTATGATGATGATCGGGGGTCAAGTGTGAAGCTTGAAGGCTGTGG 1310  
Db 961 AGGCTTTTCTTATGATGATGATCGGGGGTCAAGTGTGAAGCTTGAAGGCTGTGG 1020  
Qy 1311 CAGTTTCTTAAACAAGAGTGAAGATGCAATATAGTGAATCTTCTTCTATAGAAA 1370  
Db 1021 CAGTTTCTTAAACAAGAGTGAAGATGCAATATAGTGAATCTTCTTCTATAGAAA 1080  
Qy 1371 TTTCTCTAGTGTGATGCTTTTGTATAGCAATTTATGCAAGTAGAACTTCTTGA 1430  
Db 1081 TTTCTCTAGTGTGATGCTTTTGTATAGCAATTTATGCAAGTAGAACTTCTTGA 1140  
Qy 1431 AATTGGA-TCAATCTCTCAAACTGCTGCTTTAACAACCTTAAGTATATATAT 1489  
Db 1141 AATTGAGAGCATCTCTCAAACTGCTGCTTTAACAACCTTAAGTATATATAT 1200  
Qy 1490 CTGAATCAATGTTGCAATTTCAACAATTTTCAAGTGTCTTCAAGAGTAGAATCCA 1549  
Db 1201 CTGAATCAATGTTGCAATTTTCAACAATTTTCAAGTGTCTTCAAGAGTAGAATCCA 1260  
Qy 1550 TCTCATTTCTGAGATGGAATCTTTGCTATCATAGAGAAATTTCTCATCTGTCAA 1609  
Db 1261 TCTCATTTCTGAGATGGAATCTTTGCTATCATAGAGAAATTTCTCATCTGTCAA 1320  
Qy 1610 GTTTATATAGATGATGAGCAATACAGTATGCTTCAAGGCTCACTTCACTTTTAAT 1669  
Db 1321 GTTTATATAGATGATGAGCAATACAGTATGCTTCAAGGCTCACTTCACTTTTAAT 1380  
Qy 1670 CCAAGTCTCTGCTGCTTCTAACAACAATGAGTCTTCTTCAATGAGCTTGAAGC 1729  
Db 1381 CCAAGTCTCTGCTGCTTCTAACAACAATGAGTCTTCTTCAATGAGCTTGAAGC 1440  
Qy 1730 TCTCCAAATGATCATGAGGCTTGAATGCACTTCTTCAAAATCTCTGTAATATTTATA 1789  
Db 1441 TCTCCAAATGATCATGAGGCTTGAATGCACTTCTTCAAAATCTCTGTAATATTTATA 1500

Qy	1790	TTTGA	CTCCCATG	AATCATG	AAATGTTCTT	AAAGGAC	CTGGAATG	GTGAATCTTCC	1849
Db	1501	TTTGA	CTCCCATG	AATCATG	AAATGTTCTT	AAAGGAC	CTGGAATG	GTGAATCTTCC	1560
Qy	1850	AAAAG	TTTTCAATTT	CTTAGT	CCAGATCC	ATCATCC	AGAGATCC	ACTTTCATATGCC	1909
Db	1561	AAAAG	TTTTCAATTT	CTTAGT	CCAGATCC	ATCATCC	AGAGATCC	ACTTTCATATGCC	1620
Qy	1910	AGTAT	ATACCTT	ATATGAAATG	TATTTCTT	CAATAAT	ATATATAG	CTTGAAGTTGAAATTA	1969
Db	1621	AGTAT	ATACCTT	ATATGAAATG	TATTTCTT	CAATAAT	ATATATAG	CTTGAAGTTGAAATTA	1680
Qy	1970	TTGAT	CCATTTT	CTGCCAAAA	TATGATGTGTG	2000			
Db	1681	TTGAT	CCATTTT	CTGCCAAAA	TATGATGTGTG	1711			

```

/ Patent No. 6828097
/ GENERAL INFORMATION:
/ APPLICANT: JOAN, KNOLL
/ APPLICANT: ROGAN, PETER
/ TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
/ FILE REFERENCE: 30307
/ CURRENT APPLICATION NUMBER: US/09/573,080A
/ CURRENT FILING DATE: 2000-05-16
/ NUMBER OF SEQ ID NOS: 479
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 148
/ LENGTH: 2418
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: repeat region
/ LOCATION: (1)..(2418)
/ OTHER INFORMATION: trigger1
/ PUBLICATION INFORMATION:
/ PUBLICATION INFORMATION:
/ AUTHORS: Jurka, J; Malichiewicz, J; Milosavljevic, A
/ TITLE: Prototypic sequences for human repetitive DNA
/ JOURNAL: Journal of Molecular Evolution
/ VOLUME: 35
/ ISSUE: 4
/ PAGES: 286-291
/ DATE: 1992-10-
/ DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
/ DATABASE ENTRY DATE: 1996-01-26
/ DATABASE ENTRY DATE: 1996-01-26
/ US-09-573-080A-148

```

Query Match	31.8%	Score 636.8	DB 3	Length 2418
Best Local Similarity	82.1%	Pred. No. 1e-121		
Matches 859	Conservative 1	Mismatches 133	Indels 53	Gaps 9
QY	979	CAGGATACCTTGAGATACCTGGGTTTGTTCCATACACACCATATATACAAATATG	1038	
DB	2418	CAGGATACCTCGGAGATATTGCGGGTTCGGTTCCAGACCGGCAATA-----	2370	
QY	1039	CAGAAGTCGATATTCACATTAAGTAGTCACACAACTTTTGGTTCCTCCAGTCATAT	1098	
DB	2369	---AAGCCAAATATCGCAATTAAGCCAGTCACGCAATTTTGGTTTCCAGTGCATAT	2314	
QY	1099	AAAAGTTTGTCTATACACACTGTAGTCTGTTAAGTGTGCATATGTTATGTCTAAA	1158	
DB	2313	AAAAGTTATGTTTACGCTATACGTAGTGTATTAAGTGTGCATATGTTATGTCTAAA	2254	
QY	1159	AA---ACGATATCCTTAATTTTAAATATGCTTATATCTAAATAAATGCTAAACATCTT	1214	
DB	2253	AACCAATGTATACCTTAATTTTAAATAATCTTTATTTGTCTAAATAAATGCTAAACATCATCT	2194	

QY	1215	TAGAGATTCAGAGATGTTGAATCTTTTGGCTGGTGAAGGCTTTTCTTAATGAGACGA	1274
Db	2193	GAGCTTCAGGCAATCTGAATCTTTTCTGGTGAAGGCTTTCGCTTGATGTTGATGGC	2134
QY	1275	T-----CGGGGCTCAGGTGCTGAAGCTTAGGGTGCCTGTGGCAGTTTCTTAA-	1322
Db	2133	TGCTGACTGATCAGGGTGTGGTGTGGAAGGTTGGGGTGGCTGTGGCAATVTTCTTAA	2074
QY	1333	----ACAAAGTGAAGATTGGCAATATGATGTGACTCTCTTCAATGAAATTTCTCTC	1378
Db	2073	TAAACAACAATGAAGTTTGCCGATCAATTGACTCTTCTTCAATGAAGATTTCTCTG	2014
QY	1379	TAGTGTGATGATCTTTTGTATAGATTTTATGACACAGTAGAATCTTTTGAATAATGGA-	1437
Db	2013	TAGCATGTGATGCTGTTTGGACAGCATTTTACCAAGTAGAATCTTTTCAAAATTTGAG	1954
QY	1438	TCATCTCTCTAACCCTGCTCTGCTTTACACACTAAGTTATATATTTTGAATCC	1497
Db	1953	TCATCTCTCTAACCCTGCTCTGCTTTATCACTAAGTTATATATTTCTTAAATCC	1894
QY	1498	ATTGTGTGATTTCAACAATTTTTCACAGTCTTTCACAGAGAGTAGATTCATCTCATTT	1557
Db	1893	TTGTGTGATTTCAACAATGTTTACACATCTTTCACAGAGATGATTTCAATCTCAGA	1834
QY	1558	CCTGAGATGATCTTTTGCTCATCCATTAAGAAATTTCTCATCTGTTCAAGTTTATC	1617
Db	1833	AACCACTT-----TCTTTGCTCATTCATTAAGAAACATCTTCATCCGTTCAAGTTTATC	1778
QY	1618	ATGAGATTCAGCAATTAAGTCAATGTCTTTCAGGCTCACTTCACTTTTAAATTCAGTTCT	1677
Db	1777	ATGAGATTCAGCAATTAAGTCAATCTTTCAGG-----CTCACCTTCAATTAATTCAGTTCT	1723
QY	1678	CTGTGCTTTTACACACATCTGTGGTCTCTCTCATTTGAAGTCTTGAACCTCTCCAA	1737
Db	1722	CTTGTGATTTTCCACACATCTGCAATTTACTTCTTCACTGAAGTCTTGAACCTCTCCAA	1663
QY	1738	TCATCCATGAGGGTTGAATGACTCTTCCAAATTCCTGTTAATATTAAATTTTGA--	1795
Db	1662	TCATCCATGAGGGTTGAATCAACTCTTCCAAATCTCTGTTAATGTTGATATTTTGACC	1603
QY	1796	-CCTCCCATGAATCATGAATGTTCTTAAATGGAACCTGGAATGGTGAATCTTTCCAAAG	1854
Db	1602	TCCTCCCATGAATCAGATGTCTTAAATGGATCTAGAAATGGTGAATCTTTCCAAAG	1543
QY	1855	GTTTCAATTTAATTAGTCCAGATCCATCCACAGAGATCCACTTCAATGACAGTGA	1914
Db	1542	GTTTCAATTTAATTGCTCCAGATCAT-----CAGAGAAATCATATCTATGAGGACGA	1488
QY	1915	TAGCCTTAAGAAATGATTTCTTCAATAATAAGGCTTGAAGTTGAATTTACTCTTGAT	1974
Db	1487	TAGCCTTAATAATGATTTCTTAAATATAAGACTGAAATGTCGAATTTACTCTTGAT	1428
QY	1975	CCATTTTTCGCAAAATGAATGTTTGG	2000
Db	1427	CCATGGCTGCGAAATGAATGTTTGG	1402

```

RESULT 4
US-09-949-016-14266
; Sequence 14266, Application US/09949016
; Patent No. 6812339
;
; GENERAL INFORMATION:
;
; APPLICANT: VENTER, J. Craig et al.
;
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;
; FILE REFERENCE: CL001307
;
; CURRENT APPLICATION NUMBER: US/09/949,016
;
; CURRENT FILING DATE: 2000-04-14
;
; PRIOR APPLICATION NUMBER: 60/241,755
;
; PRIOR FILING DATE: 2000-10-20
;
; PRIOR APPLICATION NUMBER: 60/237,768
;
; PRIOR FILING DATE: 2000-10-03
;
; PRIOR APPLICATION NUMBER: 60/231,498
;

```

PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FaastSeq for Windows Version 4.0  
SEQ ID NO: 14266  
LENGTH: 26416  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-14266

Query Match 24.2%; Score 483.8; DB 3; Length 26416;  
Best Local Similarity 74.6%; Pred. No. 4,4e-90;  
Matches 769; Conservative 0; Mismatches 192; Indels 70; Gaps 10;

978 ACAGGATCTCTGAGATACGTGGTGGTTCATCCACCAATTAATTAATAT 1037  
13527 ACAGTACACACTGACATATTTCAAGTTCAGACCACTCAATTAAGCA- 13582  
1038 GCAAGAGTGAATATCAATTAAGTGAAGTCAACAAGTCTTGGCTCCAGTCA 1097  
13583 -ATATCAATTAAGCAAGTCAATGAATTTGGTTCTCGATGAT 13631  
1098 TAAAGTTTCTTAATTAATCACTGATGCTGTAAGTGAATGTTATGCTTAA 1157  
13632 TAAAGTTTCTTAAT- -AGGCAAGCAATGATTAATGCTGAA 13674  
1158 AAAACATACCTTAATTTAAATGCTTAATTAATTAATTAATTAATTAAT 1217  
13675 AAAGTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 13734  
1218 CATTCAGTGAATGATCTTTTGGTGGAGAGTCTTTTCTTAATGATGAT- 1275  
13735 CTTTAAAGAGTGAATCTTTTGGTGGAGAGTCTTTTCTTAATGATGAT- 13794  
1276 -CGGGGTCAGGTGCTGAAGCTTAAGTGGCTGTTGAGCTTTT- -AA 1321  
13795 TGAGTATCAGGAGTGGTGGTGGTGAAGGTTGGGCTGAGCAATTTCTTCA 13854  
1322 AACACAGTGAATGATTAATCACTGATCTTCTTCAATGAAGATTTCTCTG 1381  
13855 GACACAAATGAAGTGGCACAATGATTAATCTTCTTCAAAAGATTAATCT 13914  
1382 TGTGATGCTTTTGAATGATTAATCAAGTGAAGTCTTTGAATTTGA- 1440  
13915 CATGATGCTGTTGATGATGAT- -ACAGTGAAGTCTTTCAAAATTTGA 13970  
1441 ATCTCTCAACCTGCTGCTTGAACCTTAATTAATTAATTAATTAATTA 1500  
13971 GTCCCTCAATCTTGCACAAAGCTTTAATTAATTAATTAATTAATTA 14030  
1501 GTTGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1560  
14031 GTTGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 14090  
1561 GAGATGAATCTTGTCTCAATTAAGAAATTCATCTGTTCAAGTTTATCA 1620  
14091 CACTT- -TCTTGTCAATTAAGAAATTCATCTGTTCAAGTTTATCA 14146  
1621 AGATGAGCAATTAAGTATCTTCAAGCTCACTTCAATTAATTAATTAAT 1680  
14147 AATGAGCAATTAAGTATCTTCAAGCTCACTTCAATTAATTAATTAAT 14201  
1681 GGTGTTTCAATTAAGTATCTTCAAGCTCACTTCAATTAATTAATTAAT 1740  
14202 GGTGTTTCAATTAAGTATCTTCAAGCTCACTTCAATTAATTAATTAAT 14261  
1741 TCCATGAGGTTGAATGATCTTCAAAATCTCTTAATTAATTAATTAATTA 1797  
14262 TCCATGAGGTTGAATGATCTTCAAAATCTCTTAATTAATTAATTAATTA 14321  
1798 TCCATGAGGTTGAATGATCTTCAAAATCTCTTAATTAATTAATTAATTA 1857  
14322 TCCATGAGGTTGAATGATCTTCAAAATCTCTTAATTAATTAATTAATTA 14381

1858 TTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1917  
14382 TTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 14436  
1918 CTTTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1977  
14437 TCTTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 14496  
1978 TTTTCTGCAA 1988  
14497 TGGGCTGTA 14507

RESULT 5  
US-09-949-016-17313/c  
Sequence 17313, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FaastSeq for Windows Version 4.0  
SEQ ID NO 17313  
LENGTH: 119762  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-17313

Query Match 24.0%; Score 480; DB 3; Length 119762;  
Best Local Similarity 73.9%; Pred. No. 3.7e-89;  
Matches 776; Conservative 0; Mismatches 220; Indels 54; Gaps 11;

971 AGATGATCAGGATCACTTGAAGTATCTGTTGGTTCATTAACCAACCAATTA 1030  
100972 ACAAATACAGGACCTCAAAAGTAAATGAGTTTCAATTAATTAATTAAT 100913  
1031 CAATATGCAAGATGATATCAATTAATTAATTAATTAATTAATTAATTA 1090  
100912 TGAATAT- -TGTATTAAGCAATTTGTATCACTTAATTTGTTTCCCA 100863  
1091 GTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1150  
100862 GTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 100803  
1151 GTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1210  
100802 ATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 100748  
1211 ATTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1270  
100747 ATCTGATCTTGAATGATTAATTAATTAATTAATTAATTAATTAATTA 100688  
1271 CTGAT- -CGGGGTCAGGTGCTGAAGCTTAAGGTTGCTGAGCTTTCTT 1319  
100687 TGGGCTGATGATCAGGTTGCTGCTGAAGGTTGCTGCTGCTGCTGCTT 100628  
1320 AAA- -ACAAAGTGAAGTGAATTAATTAATTAATTAATTAATTAATTA 1374  
100627 AAAAGAGCAACAGAGTTTGCACATTAATTAATTAATTAATTAATTA 100568  
1375 TCTTGAATGATGATGATTTTGAATTAATTAATTAATTAATTAATTAATTA 1434  
100567 TCTTGAATGATGATGATTTTGAATTAATTAATTAATTAATTAATTAATTA 100508

QY 1435 GGA-TCAATCTCTGCAAAACCTGCTGTGTTTAAACAACCTAAGTAAATATATTCGA 1493  
DB 100507 GGAGTCATCTCTCCCAACCCCTGCTACGCTTATCACTAAGTATATATTCGA 100448  
QY 1494 ATCCATTTGTTGATTTTCAACAATTTTCAAGTGTCTTCAACAAGATATGATTCATCTC 1553  
DB 100447 ATCCCTCTGTTGATTTTCAACAATTTTCAAGTGTCTTCAACAAGATATGATTCATCTC 100388  
QY 1534 ATTTCTGAGATGATGATCTTGTCTCATATCATATAGAAATATCTCATCTGTTCAAGTT 1613  
DB 100387 AAGAAATACGT-TCTTTGACATTCATATAGAAAGCACTCCCATTTATTAAGTTT 100332  
QY 1614 TATCATGATTTGACAGCATATACATCATGTCTTCAAGGCTCACTTCACTTTAATTCAG 1673  
DB 100331 TATCATGATTTGACAGCATATACATCATGTCTTCAAGGCTCACTTCACTTTAATTCAG 100277  
QY 1674 TTCTCTGTGTTTCTTACCAACATCTGTGTCTTCTTCTCATTTGAAGTCTTCACTCTC 1733  
DB 100276 TTCTCTGTGTTTCTTACCAACATCTGTGTCTTCTTCTCATTTGAAGTCTTCACTCTC 100225  
QY 1734 CAATCATCATGAGGTTGAGATGATGATCTTCTTCAAAATTCCTGTTAATTTATATTTT 1793  
DB 100224 AAAGTCATCATGAGGATTTGATCAATCACTTCTTCAAACTCCTGTTAATTTGATTTT 100165  
QY 1794 GA---CTTCCCATGATCATGATGATTTCTTAAATGACCTGGAATGTAATCTTTTCA 1850  
DB 100164 GACATCTCTTCCATGATCATGATGATTTCTTAAATGACCTGGAATGTAATCTTTTCA 100105  
QY 1851 AAAGTTTCAATTTTACTTATGTCATGATCATCATTCAGAGATTCATTTCAATGCCA 1910  
DB 100104 GAAGTTTCAATTTTACTTATGTCATGATCATCATTCAGAGATTCATTTCAATGCCA 100047  
QY 1911 GTTATAGCTTATGAGATGATTTTCTTCAATATTAAGGCTTGAATGTAATTTACTCTT 1970  
DB 100046 GTTATAGCTTATGAGATGATTTTCTTCAATATTAAGGCTTGAATGTAATTTACTCTT 99987  
QY 1971 TGATCATTTTCTGCAAAATAGATGTTGTG 2000  
DB 99986 TGATCATGAGGCTGAGATGATGATTTGTG 99957

RESULT 6  
US-09-949-016-14273/C  
; Sequence 14273, Application us/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14273  
; LENGTH: 237510  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(237510)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14273

Query Match 23.6%; Score 472.4; DB 3; Length 237510;  
Best Local Similarity 74.1%; Pred. No. 1.5e-87;

Matches 782; Conservative 0; Mismatches 211; Indels 63; Gaps 12;  
QY 974 TAGTACAGGCAATACCTTGAGATATCTGTGGTTGGTTCCATACCAACCAATATACAA 1033  
DB 40017 TTGTACAGGCAATACCTTGAGATATCTGTGGTTGGTTCCATACCAACCAATATACAA 39961  
QY 1034 ATATGCAAGAAAGTGATATCATATTAAGTATGATCAACAAAGTCTTTGGCTTCCAGT 1093  
DB 39960 -----TGAAATTTGCAATTAAGTATGATCAACAAAGTCTTTGGCTTCCAGT 39914  
QY 1094 CATATTAAGTTTGGCTTATATCTACATGTATGTCTGTATAGTGTCAATATGTTATGTC 1153  
DB 39913 CATATTAAGTTTGGCTTATATCTACATGTATGTCTGTATAGTGTCAATATGTTATGTC 39854  
QY 1154 TAAA-----AAAACATACCTTAATTTAAATGCTTTATTAAT-AAAAATGCTAACAA 1208  
DB 39853 TAAAATCAATGATCATATATTAATTAATTAACATTTATATGCTTAAAAATGCTAACAA 39794  
QY 1209 TCATTTGACATTCAGTATGTTTAACTTTTGTGTGTGGAAGGCTTTTCTTATGAT 1268  
DB 39793 TCATCTGAGCCTTCAGCAAGTCATATATC-TTTTGAGGTGAGAGGCTTGTCTGTGT 39735  
QY 1269 GACGT-----ATCGGGGTCAGGTGCTGAAGCTTAGAGGTGCTGTG 1310  
DB 39734 GAAGCTGATGATGATGCTGATTAATTAAGGTGTGTGTGTCTTAAAGTTGGGGTGTGTG 39675  
QY 1311 CAGTTTCTTAA-----ACAACATGAAAGATTCGAATATCAAGTATGATCTTCTTTCATG 1365  
DB 39674 CAGTTTCTTAAATTAAGATTAATTAATTAAGTTTGTGATCAACCAATCTTCTTTCATG 39615  
QY 1366 AAAGATTTCTCTTATGTGTGTATGCTTTTGTATGATTTTATGACAGATGAACTTC- 1424  
DB 39614 AAAGCTTCTCTGACCTGTGATATCTGTGTATTAACATTTTAAACCAAGCCAACTCT 39555  
QY 1425 TTTGAAATTTGATGATATCTCTTCAACCTGCTGCTTTTAAACAACCTTAATTAATA 1484  
DB 39554 TTTGAAAGGACAGATTAATCTGTCTCAACCTGCTGCTTTTAAACAACCTTAATTAATA 39435  
QY 1485 ATATTCGATATCATATGTTGTATTTCAACAATTTTCAAGTGTCTTCAACGAGATGA 1544  
DB 39435 ATATTCGATATCATATGTTGTATTTCAACAATTTTCAAGTGTCTTCAACGAGATGA 39435  
QY 1545 TTTCATCTCATTTCTGAGATGAAATCTTGTCTCATTCATATGAAGAAATTCCTCATCTG 1604  
DB 39434 TTTCATCTCATTTCTGAGATGAAATCTTGTCTCATTCATATGAAGAAATTCCTCATCTG 39379  
QY 1605 TTCAAGTTTATCATGAGATGTCAGCAATATACATGATGCTTCAAGGCTTCACTTCACTT 1664  
DB 39378 TTCAAGTTTATCATGAGATGTCAGCAATATACATGATGCTTCAAGGCTTCACTTCACTT 39324  
QY 1665 TAATTCAGTCTCTGTGTGTTTCTACCAATCTGTGTGTTCTTCTTCAATGAAGTCTT 1724  
DB 39323 TAATTCAGTCTCTGTGTGTTTCTACCAATCTGTGTGTTCTTCTTCAATGAAGTCTT 39267  
QY 1725 GAACCTTCCAGATCATATGAGGTTGGAATTCGATTTCTTCAAAATTCCTGTTAAT 1784  
DB 39266 GAACCTTCCAGATCATATGAGGTTGGAATTCGATTTCTTCAAAATTCCTGTTAAT 39207  
QY 1785 TTAATTTTGAACCTTCCAGATCATATGATGTTCTTAATGAGACCGGATGATGATTC 1844  
DB 39206 TTAATTTTGAACCTTCCAGATCATATGATGTTCTTAATGAGATGATGATGATGATTC 39147  
QY 1845 TTTCCAAAAGTTTCAATTTTACTTATGTCAGATTCATTCAGAGATCCACTTTTCA 1904  
DB 39146 TTTCCAAAAGTTTCAATTTTACTTATGTCAGATTCATTCAGAGATTCATTCAGAGATTC 39092  
QY 1905 ATGCCAGTTATAGCTTATGATGATGATTTTCTCAATTAATTAAGCTTGAAGTTGAAT 1964  
DB 39091 ATGCCAGTTATAGCTTATGATGATGATTTTCTCAATTAATTAAGCTTGAAGTTGAAT 39032  
QY 1965 ACTCTTGAATCATTTTCTGCAAAATAGATGTTGTG 2000  
DB 39031 ACTCTTGAATCATTTTCTGCAAAATAGATGTTGTG 38996

RESULT 7  
US-09-949-016-13173/C  
Sequence 13173, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13173  
LENGTH: 360470  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-13173

Query Match 23.4%; Score 468.2; DB 3; Length 360470;  
Best Local Similarity 72.6%; Pred. No. 1.2e-86;  
Matches 762; Conservative 0; Mismatches 233; Indels 54; Gaps 10;

QY 975 AGTACGCGATCCTTGAGATPCTGTGTTGGTTCCATACACCAATTAATACAA 1034  
DB 253971 ATTACGGGATVCTTAAAGATATTCAGGGTATGTTCAAGACTACCGAATA----- 253919  
QY 1035 TATGCAAGAGGATATATCAATATAAGTACACAAAGCTTTGGCTTCCAGTGC 1094  
DB 253918 -----AATCAGATATGCACTACAGATTAACCAATTTCTGGTGAACCGATGC 253867  
QY 1095 AATAAAGTTTGTCTTAATACACTGTAGTCTGTTAAGTGAATAGTATATGT 1154  
DB 253866 ATGTAAGATTAATGTTTACGCTATCTGATCTAATAGTGAATAGTATATGTGA 253807  
QY 1155 AAAAAACATTA-----CTTAATTTTAAATGCTTTTATTAATAAATGCTAACAT 1209  
DB 253806 AATAAACAATGATACATACCTTAATTAATAAATACTTTATGCTAATAAATGCTAACAT 253747  
QY 1210 CATTTGAGATTCAGAGATGTTGAATCTTTTGTGCTGAGAGGCTTTTCTTATGATG 1269  
DB 253746 TATCTGAGCTTCAGGAGATTTGAAGCTTTTGTGCTGAGAGGCTTTGCTCCATGTTG 253687  
QY 1270 ACTGAT-----CGGGGGTCAGGTGCTGAAGCTTAAGGTGCTGTGGCAGTTTCT 1318  
DB 253686 ATGGCTGCTGACGATCTGGGTGGTGTGCTGTGCTGTGGGTGCTGTGGCAATTTCT 253627  
QY 1319 TAAA-----ACAAAGTGAAGATTTGCAATACATGTTGACTCTTCTTATGAAGATT 1373  
DB 253626 TAAATATGACATATATGAAGTTGTATACATGATGACCTCTTCTCATGAAGATAT 253567  
QY 1374 CTCTAGTGTGATGCTTTTGTATGAGATTTTATGACAGTGAAGCTTTTGAAGAT 1433  
DB 253566 CTCTGATGATGATGATGCTGTGTGAAGAGCTTTAATCCAGTGAAGCTTTTGAAGAT 253507  
QY 1434 TGAATCTCTCAAAACCTGCTCTGCTTTAACAACCTAAGTTAATATATATCTG 1492  
DB 253506 TGAAGTCAATCTCTTAATCTGTGCTACTTACCACTAAGTTATGATATATCTGA 253447  
QY 1493 AATCCATTTTGTATTTCAACAATTTTCAAGATGCTTACAGAGAGTATGATCT 1552  
DB 253446 AATCTGTGTGCTATTTCAACAATGCTTACAGCAATTTTATCAGTATGATTCATC- 253388  
QY 1553 CATTTCTGAGATGATCTTTGCTCATCATTAAGAAATTTCTCATCTGTTCAGATT 1612

DB 253387 ---TTAAGAGCCCTTTCTTACTATCATACAGTAAAGTAACTCTTAATCTGTCAAGTT 253331  
QY 1613 TTATCATGATTTGAGCAATACAGTATGCTTCAAGGCTGACTCATTTTAATTCGA 1672  
DB 253330 TCATCATGAGATTTGAGCAATACAGTATGCTTCAAGGCTGACTCATTTTAATTCGA 253276  
QY 1673 GTTCTCTGCTTTTCTACACATCTGTGTTCTCTTCTTCAATGAAGCTTGAACCTCT 1732  
DB 253275 GTTCTCTGCTTTTCTACACATCTGTGTTCTCTTCTTCAATGAAGCTTGAACCTCT 253217  
QY 1733 CCAAGTATCATGAGGTTGGAATGCACTTCTTCAAAATCTCTTAATTTATATTT 1792  
DB 253216 TAAAGTATCATGAGGTTGGAATGCACTTCTTCAAAATCTCTTCAATTTATATTT 253157  
QY 1793 --TGACCTCCCATGATCATGATGTTCTTAATGAGCACTGGAATGGAATCTTTCCA 1850  
DB 253156 GACCTCTCCGATGATCATGATGTTCTTAATGAGCACTGGAATGGAATCTTTCCA 253097  
QY 1851 AAAGTTTCAATTTACTAGTCCAGATCCATCCATCCAGAGATCCATTTCAATGCA 1910  
DB 253096 GAAAGTTTCAATTTACTAGTCCAGATCCATCCATCCAGAGATCCATTTCAATGCA 253042  
QY 1911 GTTATAGCTTATGAGATGATTTCTTCAATTAATAGGCTTGAAGTTGAATTTACTCT 1970  
DB 253041 AGTATAGCTTATGAGATGATTTCTTCAATTAATAGGCTTGAAGTTGAATTTACTCT 252982  
QY 1971 TGATCATTTTCTGCAAAATAGATGTTGT 1999  
DB 252981 TGATCATGATGATGATGATGTTGT 252953

RESULT 8  
US-09-949-016-15390  
Sequence 15390, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15390  
LENGTH: 235064  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(235064)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15390

Query Match 23.3%; Score 465.8; DB 3; Length 235064;  
Best Local Similarity 73.1%; Pred. No. 3.5e-86;  
Matches 779; Conservative 0; Mismatches 232; Indels 54; Gaps 12;

QY 972 GATAGTACGAGATACCTTGTGAGATGCTGTGCTTGTGCTCATACCAATTAATATC 1031  
DB 47772 GCTAATACGAGATACCTTGTGAGATGCTGTGCTTGTGCTCATACCAATTAATATC 47831  
QY 1032 AATATGCAAGAGTGC-----ATATCAATTAAGTGAAGTCAACAAGCTTTT 1081  
DB 47832 AATATGCAAGAGTGC-----ATATCAATTAAGTGAAGTCAACAAGCTTTT 47891  
QY 1082 GGTCTCCAGTGCATATAAAGTTTGTCTATACACTGATGCTGTAAGTGTGCA 1141



```

Db      47892 GATTCCAGGCGGATATAAGTATGTTACACATGCTGATGCTATTAAGTGTGAAA 47951
Qy      1142 TAGGTTATGTCCTAAAAA---CACATACCTTAATTTTAAAGCTTATTAATAAA 1198
Db      47952 TAGCATTAATGTCCTAAAAA---CACATACCTTAATTTTAAAGCTTATTAATAAA 48011
Qy      1199 ATGCTAACATCATTTGAGCATTCAGTGAAGTTGATCTTTTGGCTGGTGAAGGCTTTT 1258
Db      48012 ATGCTAACATCATTTGAGCATTCAGTGAAGTTGATCTTTTGGCTGGTGAAGGCTTTT 48070
Qy      1259 TCTTATGATGATGATGAGGGGGTCA-----GGTGTGAAGCTTGAAGGCTGG 1307
Db      48071 CCTCATGTCAGTGGCTGCTGATCATGATGATGGGTTACTGCAAGGTTGAGTGAATG 48130
Qy      1308 TGGCATGTTCTTAAACAA---GTGAAGTTCATATATGATGATGATCTTCTTCA 1363
Db      48131 TGACATTTCTTAAATTAAGACAAATGAAGTTTGACATATGATGATGATCTTCTTCTT 48188
Qy      1364 TGAAGATTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1423
Db      48189 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 48248
Qy      1424 CTTTGAATTTGGA-----TCAATCTCTCAAACTCTGCTTGAACCTTAAGT 1478
Db      48249 CTTTGAATTTGGA-----TCAATCTCTCAAACTCTGCTTGAACCTTAAGT 48308
Qy      1479 AATATAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1538
Db      48309 --TATGATGCGGTAAATTTTGTGTGATTCACAGATGATGATGATGATGATGATGAT 48366
Qy      1539 AGTAGATTCATCTCATTTCTGAGATGATGATGATGATGATGATGATGATGATGATGAT 1598
Db      48367 AGTAGATTCATCTCATTTCTGAGATGATGATGATGATGATGATGATGATGATGATGAT 48422
Qy      1599 CATCTGTTCAATTTTATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1658
Db      48423 CATCTGTTCAATTTTATCATGATGATGATGATGATGATGATGATGATGATGATGAT 48477
Qy      1659 CACTTTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1718
Db      48478 CACTTTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 48537
Qy      1719 AGCTTTGAACCTCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1778
Db      48538 AGCTTTGAACCTCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 48597
Qy      1779 TAAATTTAATTTTGA---CCTCCATGATGATGATGATGATGATGATGATGATGAT 1835
Db      48598 TAAATTTAATTTTGA---CCTCCATGATGATGATGATGATGATGATGATGATGAT 48657
Qy      1836 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1895
Db      48658 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 48713
Qy      1896 CCACCTTCAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1955
Db      48714 CCACCTTCAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 48773
Qy      1956 GTTGAATTTACTCTTGAATTCATTTTCTGCAAAATGATGATGATGATGATGATGAT 2000
Db      48774 GTTGAATTTACTCTTGAATTCATTTTCTGCAAAATGATGATGATGATGATGATGAT 48818

```

```

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 17422
; LENGTH: 164061
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17422

Query Match      23.3% Score 465.6; DB 3; Length 164061;
Beet Local Similarity 74.9%; Pred. No. 3.5e-86;
Matches 789; Conservative 0; Mismatches 204; Indels 61; Gaps 14;

Qy      975 AGTACAGGATACCTTGGAGATGATGATGATGATGATGATGATGATGATGATGAT 1034
Db      75766 ATTATAGGATATCTTGGAGATGATGATGATGATGATGATGATGATGATGATGAT 75817
Qy      1035 TATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1094
Db      75818 -----GAGTGAATGACAAATTAAGCAAGTCAAAATTTATTTCTTTGCAATGTC 75870
Qy      1095 AATATAAGTTTGTGTTATPACACCTGATGATGATGATGATGATGATGATGATGATGAT 1150
Db      75871 AATATAAGTTTGTGTTATPACACCTGATGATGATGATGATGATGATGATGATGATGAT 75930
Qy      1151 GTCTAAAAAACAATACCTTAA--TTTAAATGCTTTATTAAC--TAAAAATGCTTAACA 1208
Db      75931 TAAAAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 75990
Qy      1209 TCATTTGACATTCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1265
Db      75991 TCATTTGACATTCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 76050
Qy      1266 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
Db      76051 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 76110
Qy      1318 TTAATA-----ACAAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 1372
Db      76111 TTAATAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 76170
Qy      1373 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1432
Db      76171 TATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 76230
Qy      1433 TTGGA--TCAATCTCTCAAACTCTGCTGCTTAAACCTTAAGTAAATATATATCT 1491
Db      76231 TTGGAATCAATCTCTCAAACTCTGCTGCTTAAACCTTAAGTAAATATATATCT 76290
Qy      1492 GAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1551
Db      76291 AATCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 76350
Qy      1552 TCATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1611
Db      76351 TCATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 76406
Qy      1612 TTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1671
Db      76407 TTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 76461
Qy      1672 AGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1729
Db      76462 AGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 76521
Qy      1730 TCTCCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1789

```

```

RESULT 9
US-09-949-016-17422
; Sequence 17422, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307

```



DB 76522 TTCCCATCATCAGAGATGGAATCACTTCTCCAACTGTTTATGTGANA 76581  
QY 1790 TTTTGA---CTCCCATGATCATGATGTTCTTANAGCACTGGAATGCTAACTT 1846  
DB 76582 TTTTGTCTCTTCCCATGATCATGATGTTCTTANAGCATTAAGATGGAATCTT 76641  
QY 1847 TCCAAAGGTTTCAATTTACTAGTCAGATCATCATCCAGAGATCACTTCAAT 1906  
DB 76642 TCCAGAGGTTTCAATTTACTTGGCCAGATTCAT-----CAGAGATCACTGTCTGT 76696  
QY 1907 GCCAGTTATAGCCTTATGAGATGATTTCTTCAATATAAG---GCTGAAAGTTGANA 1962  
DB 76697 GGCATATATAGCCTTACAGATGTTGTTCTTAATATATAAGATTTCTGAAAGTCTAAA 76756  
QY 1963 TTAATCTTATCATCTTTCTGCAAAATAGATCT 1996  
DB 76757 TGACTCTTATCATAGGCTACAAATAGATAT 76790

## RESULT 10

US-09-949-016-11869/c  
; Sequence 11869, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 11869  
; LENGTH: 194714  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11869

Query Match 23.1%; Score 462.6; DB 3; Length 194714;  
Best Local Similarity 74.3%; Pied. No. 1.5e-85;  
Matches 778; Conservative 0; Mismatches 209; Indels 60; Gaps 13;

QY 979 CAGGCATCTGAGATGATCTGGGTTGGTTCCCATACACACATAATACAAATATG 1038  
DB 62690 CAGGTTACTAGAGATATATACAGGTTGGTTCTAGACCACACATAAAGCAATAT- 62632  
QY 1039 CAGAAGTGGATATCAATAAAGTATGATCACAAGATCTTTGGCTCCAGTGCAT 1098  
DB 62631 -----TATGATATAAAGT-----CACACAAATTTTGGTTCCAGTGCAT 62589  
QY 1099 AAAAGTTTCTTATATCACTAGTCTGTTAAAGTGCATATGTTATGCTAA- 1157  
DB 62588 AAAAGTTATATTAATCTAGTCTGTTAAAGTGCATATGCTAA- 62529  
QY 1158 -----AAAACATCTTAAATTTTAAAGCTTATATTAATAAATGCTAACATCAT 1213  
DB 62528 TCACATTAATATATCTTAAATTTTAAATATCTTGTCTAGAAAATGCTAAGATCATC 62469  
QY 1214 TGAGCATTCAGTATGTTATCTTTTGTCTGAGAGGATCTTTTCTTATGATGACTG 1273  
DB 62468 TGAGCTCCAGTATGTTATCTTTTGTCTGAGAGATCTCGCTCCAGCTTATGAG 62409  
QY 1274 AT-----CGGGGCTCAGGCTGAGAGTATAGGTTGCTGCGAGTTTCT----- 1318  
DB 62408 CTGCTAGTATGATCAGGATGATGTTGCTGAGAAATGAGGATGATGAGGATTTCTTAA 62349

QY 1319 -TAAACACAGTGAAGATGGAATATACGTTGACTCTTCTCTTCAATGAAGATTTCTCT 1377  
DB 62348 ATTAACACACATGAAGTTTGGCAACAATATGACTCTTCTTCAATGAAGATTTCTCT 62289  
QY 1378 CTAGTGTGATGCTTTTGTATGCAATTTTATGACAGATGAACCTTTTGAATTTGA 1437  
DB 62288 GTAGCATGCTACATTAATTTGATGATGATTTTACCCACMACAGAACCTTTCAAAATTTGA 62229  
QY 1438 -TCAATCTCTGAACCCGCTCTGCTTTAACAAGTATGTTAATATATTTCTGATC 1496  
DB 62228 GGCATCTCTGAACCCGCTCTGCTTTAACAAGTATGTTAATATATTTCTAAT 62169  
QY 1497 CATTTGTCATTTTCAACATTTTCAACAGTCTTCCACAGAGATGATTTCAATCTCAT 1556  
DB 62168 CTTTTCGTTATTTGACATGTT-ATGSCATCTTCATATGGAATATTTATCTCAAG 62110  
QY 1557 TCCATGATGATCTTTGCTCATTCATTAAGAAATTTCTCATCTGTTCAAGTTTAT 1616  
DB 62109 AAAACACTT-----TCTTTCATCTGTAAGATCACTCTCATTTCTCAAGTTTAT 62054  
QY 1617 CATGATGTCAGCAATATAGCATGCTTCAAGGCTCATCTTCAATTTATTTCCAGTTC 1676  
DB 62053 CATGATGTCAGCAATATAGCATGCTTCAAGGCTCATCTTCAATTTATTTCCAGTTC 62000  
QY 1677 TCTTGTCTTCTACACATCTGTGTTCTCTTCCATTTGAAGTCTTCAACCTTCCAA 1736  
DB 61999 TCAATCTTCTTCCACACATCTGTGTTCTCTTCCATTTGAAGTCTTCAACCTTCCAA 61940  
QY 1737 GTCATTCATGAGGTTGATGATGATCTTCCAAATTCCTGTTAATATTTATTTTGC 1796  
DB 61939 GTCATTCATGAGGCTGATGATGATCTTCCAAATTCCTGTTAATATTTTGC 61880  
QY 1797 CTC---CATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1853  
DB 61879 CTCCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 61820  
QY 1854 GGTTCATATTTACTTATGTCAGATCATCATCCATCCAGAGATCCATTCATGACGAT 1913  
DB 61819 GG-TTTCATTTTACTTGTGCTTATCATCT-----CAGAGATCATTTATATGCGACT 61766  
QY 1914 ATAGCCTATGATATGATTTTCTTCAATATATAGGCTTGAAGTTGAATTAATCTCTTGA 1973  
DB 61765 ATAGCCTATGATATGATTTTCTTCAATATATAGGCTTGAAGTTGAATTTTCTCTGA 61706

QY 1974 TCCATTTTCTGCAAAATAGATGTTG 2000  
DB 61705 TCCATAGGCTGTTGAATGATATTTG 61679

## RESULT 11

US-09-949-016-15474/c  
; Sequence 15474, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 15474  
; LENGTH: 196714  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15474

Query Match 23.1%; Score 462.6; DB 3; Length 196714;  
Best Local Similarity 74.3%; Pred. No. 1.5e-85;  
Matches 778; Conservative 0; Mismatches 209; Indels 60; Gaps 13;

```
OY 979 CAGGATACCTTGGAGATGCTGGGTTGGTTCCATACACCAATATATCAATATG 1038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112690 CAGGATACCTTGGAGATGCTGGGTTGGTTCCATACACCAATATATCAATATG 112632
OY 1039 CAGGATACCTTGGAGATGCTGGGTTGGTTCCATACACCAATATATCAATATG 1098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112631 -----TATATATATATATATATATATATATATATATATATATATAT 112589
OY 1099 AAAAGTTTGGCTTATATACCTGATGCTGTTAAGTGTGCAATAGTGTATATGCTAA- 1157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112588 AAAAGTTTATTTATATATATATATATATATATATATATATATATATATAT 112529
OY 1158 ----AAACATACCTTATATATATATATATATATATATATATATATATATAT 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112528 TCACATATCATATATATATATATATATATATATATATATATATATATATAT 112469
OY 1214 TGAGATTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112468 TGAGCTCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 112409
OY 1274 AT-----CGGGGGTCAAGTGTGATGATGATGATGATGATGATGATGATGAT 1318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112408 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 112349
OY 1319 -TAAACACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112348 ATAAACACAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 112289
OY 1378 CTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112288 GTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 112229
OY 1438 -TCAATCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112228 GGCATCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 112169
OY 1497 CATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112168 CTTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 112110
OY 1557 TCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112109 AAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 112054
OY 1617 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112053 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 112000
OY 1677 TCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111999 TCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 111940
OY 1737 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111939 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 111880
OY 1797 CTC---GCATGATGATGATGATGATGATGATGATGATGATGATGATG 1853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111879 CTCCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 111820
OY 1854 GGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111819 GG-TTCATGATGATGATGATGATGATGATGATGATGATGATGATG 111766
OY 1914 ATAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111765 ATAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 111706
```

```
OY 1974 TCCATTTTCGCAAAATATGATGTTGTG 2000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111705 TCCATGAGCTGTGATGATGATGATGTTGTG 111679
```

```
RESULT 12
US-09-949-016-14179
; Sequence 14179, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14179
; LENGTH: 340380
; TYPE: DNA
; ORGANISM: Human
; PEPTIDE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(340380)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14179
```

Query Match 23.1%; Score 461.2; DB 3; Length 340380;  
Best Local Similarity 73.2%; Pred. No. 3.3e-85;  
Matches 779; Conservative 0; Mismatches 218; Indels 67; Gaps 12;

```
OY 956 TTTCTGTAAGGCTAGATAGTACAGGATACCTTGGAGATGATGAGTGGTGGTTCAT 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34227 TTTATTTATGCTTTTCTAGTACAGGATATCTTGGAGTTATTCAGGCTCAGTTCCG 34286
OY 1016 ACCACCAATATATCAAAATATGCAAGATGATATCAATTAAGTACATCAAG 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34287 ATCCTGCAATA-----AAGTAATATCAATTAAGCAAGTACACACA 34331
OY 1076 TCTTTTGTCCGAGTATATTAAGTTTGTCTTATACATGATGATGATGATG 1135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34332 TTTTGTGTTCTTATGATGATGATGATGATGATGATGATGATGATGATG 34390
OY 1136 GTGCAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34391 GAGCAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 34450
OY 1194 AAAAATGCTAATCAATTTGAGATGATGATGATGATGATGATGATGATGATG 1253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34451 -AAAATGCTAATCAATGATGATGATGATGATGATGATGATGATGATGATG 34509
OY 1254 TCTTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34510 TCTTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 34569
OY 1303 GGTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34570 GACTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 34629
OY 1361 TCATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34630 TCGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATG 34689
OY 1421 CTTCTTTGAAATGGA-TCAATCTCTCAAACTGCTGCTTATCAACTAAGTTA 1479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34690 CTTCTTTCAAAATGGAATGATGATGATGATGATGATGATGATGATGATG 34749
```

OY	1480	ATATATAATTCGATCATCTGTGTGATTTAAACAATTTTCAAGTGCTTACACAGA	1539
Db	34750	ATATATATTTCTAAATCACTTTGTGTATTTCAACAGTGTCACAGATCTTCATACAA	34809
OY	1540	GTAGATTCATCTCATTTCTCGAGATGGAATCTTTGCTCATTCATAGAAGAAATTCCTC	1599
Db	34810	GTAGATTCATCTCAAGAAACACTT-----TCTTTGCTTACCATAGGAAGCACTGCTC	34865
OY	1600	ATCTGTTCAAGTTTATTCATGATGATTCAGCAATATACGTCATGCTTCAGGCTCACTTC	1659
Db	34866	ATCGGTTCAAGTGATGATCATGATGATTCAGCAATATGATGATGCTTCAGGCTCACTTC	34925
OY	1660	ACTTTTAATTCAGATGCTCTTGCTGCTTTCTACCAACATCTGTGCTCTCTTCATTTGA	1719
Db	34926	TAAATTTCTA-----GTTGTTCTTTCCACCACTGAGGTGATCTTCTTCATTTGA	34975
OY	1720	GTCCTTGAACCTCTCCAGATCATCATAGAGGTTGGAATGACCTTCTTCCAAATTCCTGTT	1779
Db	34976	GTCCTTGAACCCCTCAAGTCACTCTTGAGTTTGAAGAACACTTCTTCCAAAGCTCTGTT	35035
OY	1780	AATATTAATATTTTGA---CCTCCATGATTCATGAATGTTCTTAAATGGCACTGGAAAG	1836
Db	35036	AATATTAATATTTTCACTCTCTCCCATGAAATCACAAAATGTTCTTAATGGCATTTGGAAAT	35095
OY	1837	GTGAATCCCTTCCAAAAGTTTTCAAATTTTACTTAAGTCCAGATCATCATCCAGAGATC	1896
Db	35096	GTGACCCCTTTC-----ATTGACTTTGACCAAGCACTC-----TAGAGGAAAT	35138
OY	1887	CACCTTCAATGCCAGTTATATAGCTTATAGGAATGATATTTCTTCAATATATAGGCTTGAAG	1956
Db	35139	AATATATATATGGCAGCATATAGCATCACAAAATCATATCTCCGTATATATAGACATGAAG	35198
OY	1957	TTGAAATTAATCCTTGATTCATTTCTCGAAAATAGAATGTGTGG	2000
Db	35199	TTAAATATGACTCTTGAACCATAGGCTGCAAAATAGATGTGCAG	35242

```

RESULT 13
US-09-949-016-13358
; Sequence 13358, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13358
; LENGTH: 260247
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13358

```

Query Match	22.7%	Score	454.2	DB 3	Length	260247	
Best Local Similarity	73.6%	Pred. No.	8.6e-84				
Matches	777	Conservative	0	Mismatches	215	Indels	65
						Gaps	13
QY	970	TAGATAGTACAGGCATACCTTGGAGATACCTGTGGGTTGGTTCATACACCACACATATAT	10259				
Db	165186	TAGAAGATACAGGCATACCTTGGAGATATTCGAGGTTTAGTCCAGACACACACATAT--	16523				
QY	1030	ACAAATATGCAGAGTGGATTCATCAATAAAGTGAAGTCACACAGTCTTTGGCTTCCC	10839				

Db	165244	-----AAGTAATATCAAAATAAAGTAATCAACAACCTTTTGTGTTCC	165290
Oy	1090	AGTGCATATAAAGTTTGGCTTATCTACATCTATAGTCTGTTAAGTGCATATAGTCTTA	1149
Db	165291	ACTACATATAAAGGTATGTTTACCTATACGTATGTCTTTAAAGTGTGCATATGACATTA	165350
Oy	1150	TGTCATAAAAAACACATA-----CCTAATTTTAAATGCTTTATTTACTATAAAAAATGCTA	1204
Db	165351	GGTCTAAAAAAACCATATGCATACCTTAATTTAAATTAATCTTATTTGCTAAAAAAATTTTA	165410
Oy	1205	ACAAATCATTTGAGCAATTCAGTAGTGTATCTTTTGTGTGTGAAGAGTCTTTTCTTAT	1264
Db	165411	GCAATCATCTAAGCTTGCAGCAAGTTAATATCTTATGCTGTGGGAGGGTTTTGACTCAA	165470
Oy	1265	TGATG-----ACTGATCGGGGGGTCAAGTGTCAAGCTTTAGGGTGGCTGTGCAGT	1314
Db	165471	TATGATGAGCTGCCCACTGAATCAGGATATGGTGTCTGAAGGTGGTGGTGGCCAT	165530
Oy	1315	TTCT-----TAAACAACAGTGAAGATTTGCAATATCAGTTGACTCTTCTTCATGAAG	1369
Db	165531	TTTTTTAATTAAGACAAATGAAGTTTGCACATCAATTAAGCTCTTCTTTCAGAAAAG	165590
Oy	1370	ATTTCTCTATGATGTGTGATGCTTTTGTATAGCATTTTATGACAGATGAACCTTCTTGA	1429
Db	165591	ATTTCTGACGACATGTGTGTGTGTATGTTATTTTAAACCAAGTGAACCTCTTTCA	165650
Oy	1430	AAATTTGGA-TCATATCTTCAAAACCTGTCTGTCTTTAACAACCTAAGTTATATATAT	1488
Db	165651	AAATTTGAGTAGCTTCTCAGACCTGTCTGTCTTTATCAACATGAAGTTTATATATATAT	165710
Oy	1489	TCTGAATCATTTGTTGTTCATTTCAACATTTTTCACAGTGTCTTACACGAGATAGATTC	1548
Db	165711	TCTAAGTCTTTAGTGTATTTTCAACATCTTTCCAGCATCTTCAACAGG-CTAAGATTC	165769
Oy	1549	ATCTCATTTCCAGATGGAATCTTTGCTCATCCATAAGAAAGAAATTCCTCATCTGTCCA	1608
Db	165770	ATCTC-----TAGAATATCACTTTTGCTATTAAGAAGCACTGCTATCTGTCT	165818
Oy	1609	AGTTTATCATGAGATTTGCAGCATATACGTCAATGTCTTCAGGCTCATCTTCACTTTAAT	1668
Db	165819	ACTTTTATATAAGATTGACACCAATTCATTCACATCTTCA-----TTCTTACTATATAT	165873
Oy	1669	TCCAGTTCTCTGGCGTTTCTACACACATCTGTGGTTCCTCTCCATTTGAAGCTTTGAAC	1728
Db	165874	TGTTGTTCTCTGTCTACTTCCACCAATTTTGCAGTTACTTACTTCATTTGAAGCTTTGAAA	165933
Oy	1729	CTCTTCAAGTCATCCATGAGGGTTGGAATGCACTTCTTCAAAATTCCTGTTAATATTTAT	1788
Db	165934	CCCTCAAGTCATCTGTGAGAGTTAAGATTAACTTCTTCAAACTCCTGTTAATATGAT	165993
Oy	1789	ATTTTGAAC---TCCATGAATCATGAATGTCTTAAATGCGACCTGGAATGTGAATCT	1845
Db	165994	GTTTGTGACCTCTCTAATGAATCTCGAATTTTCTTCATGACATCTAGAAATG---AAACCT	166051
Oy	1846	TTTCAAAAAGTTTCAATTTACTTATAGTCAGATCCATCCATCCAGAGATCCACTTTCAA	1905
Db	166052	TTTCAAGAGGTTTCCAATTTGTTTTG-CCAATTCAT-----AAGAAATACATATTTTA	166105
Oy	1906	TGCCAGTTATAGCCTTATGAGATGTATTTCTTCATATAAAGGCTTGAAGTTGAAATTTA	1965
Db	166106	TGACAGCTATA-CATTACACAAATGTGTTCTTTAATAATATAAGCTTGAATAATTAAGA	166164
Oy	1966	CTCTTGAATCATTTTCTGCAAAATGAATGTTTTG 2000	
Db	166165	CTCTTGAATCATGGCTGACGAATGGGTATTTG 166199	

RESULT 14  
US-09-949-016-15594/C  
; Sequence 15594, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15594  
LENGTH: 59948  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15594

Query Match 22.7%; Score 453.8; DB 3; Length 59948;  
Best Local Similarity 72.9%; Pred. No. 7,6e-84;  
Matches 772; Conservative 0; Mismatches 222; Indels 65; Gaps 12;

QY 971 AGATAGTACAGGACATCTTGAGATCTGTGGGTTGGTTCCATACCAACCAATATA 1030  
DB 53894 AAAATGTACAGGACATCTTGAGATCTGTGGGTTGGTTCCATACCAACCAATATA 53835  
QY 1031 CAAATATGCAAGATGATATACCAATATAGTGTACACAGCTTTGGCTTCCCA 1090  
DB 53834 CAAG-----TATACAAATAAATAGATCAAAA--TTTTTGTTCCTA 53792  
QY 1091 GTGCATATTAAGTTTGTCTTACTACACTGTAGCTGTGTAAGTGCATATGTTAT 1150  
DB 53791 GTGCATATTAAGTTTGTCTTACTACACTGTAGCTGTGTAAGTGCATATGTTAT 53722  
QY 1151 GTCTAAAAAAC-----ACATACCTTAATTTTAAATGCTTTATTAATAAATGCTAA 1205  
DB 53731 GTCTAAAAAACAAATATACATCTTAATTTAAACACCTTTATTCCT-AAAAATGCTAA 53673  
QY 1206 CAATCATTTGACATTCAGTGTGTATCTTTTGGTGGGGAAGCTTTTCTTAT 1265  
DB 53672 TGATCATCTCAGCTTCACAGCAAGTGGCCATCTTTTGGTGTGAGAGGCTTGTGTAT 53613  
QY 1266 GATGACTGATC-----GGGGGTGAGGTGCTGAAGCTTAGGGGTGCTGTGCAAT 1314  
DB 53612 GTTGATATCTCTGACTGATCAGAGCGGGTTCGTAAGATTTGGGTGGCTGTGCAAT 53553  
QY 1315 TTCTTAAAAACA--ACAGTGAAGATTGCAATATCAGTGTGACTCTTCCTTCATGAAGATT 1372  
DB 53552 TTCTTAAAAATAAGACAAATATGAAATTTGCCACATTTGACTCTTCCTTTATGAAGATT 53493  
QY 1373 TCTCTCTAGTGTGTATGCTTTTGTATACATTTTATGACATGACATGAACATCTTTGAAA 1432  
DB 53492 TCTCTCTAGTGTGTATGCTTTTGTATACATTTTATGACATGACATGAACATCTTTGAAA 53433  
QY 1433 TTG-GATCAATCTCTCAAAACCTGCTGCTTTAACAACCTTAATTAATATTTCT 1491  
DB 53432 ATGTTGTCAATCTCTCAAAATTTGCTGCTGCTTTATCAACATTAATTAATTTCT 53373  
QY 1492 GAATCC-----ATTGTTGTCAATTTCAACATTTTCAAGTGTCTTACCAAGAGTAGA 1544  
DB 53372 AAATCCTTATGAGTTGTGGTCAATTTTAAACATGTTCAAGACTCTTTATTAAGAGTAGA 53313  
QY 1545 TTCCATCTCATTTTCAGAGATGGAATCTTTGCTCATTCATAGAGAAATTTCTCATCTG 1604  
DB 53312 TTCCATCTCATTTTCAGAGATGGAATCTTTGCTCATTCATAGAGAAATTTCTCATCTG 53257  
QY 1605 TTCAAGTTTATCATAGATGAGATGCAAGATPACAGTGTGCTTACAGGCTTCACTTACT 1664  
DB 53256 TTCAAGTTTATCATAGATGAGATGCAAGATPACAGTGTGCTTACAGGCTTCACTTACT 53202  
QY 1665 TAATTCAGTTCTCTTGTCTTCTTACCAATCTGTGTCTTCTTCAATGAAGTCTT 1724

DB 53201 TAATTCAGTTCTCTTGTCTTCTTACCAATCTGTGTCTTCTTCAATGAAGTCTT 53142  
QY 1725 GAACCTCTCAAGTATCATCATGAGGTTGGAATGACATCTTCCAAATTTCTGTAAAT 1784  
DB 53141 AAACCCCTCAAGTATCATCATGAGGTTGGAATGACATCTTCCAAATTTCTGTAAAT 53082  
QY 1785 TTATATTTTGA--CTTCCATGATCATGATGTTCTTAATGCACTGTGAATGTGA 1841  
DB 53081 TGATATTTTGAAGCTTCTTCCATTAATCATGATGTTCTTAATGCACTGTGAATGTGA 53022  
QY 1842 TCCCTTCCAAAGTTTTCATATTTACTATGTCAGATTCATCCAGAGATCTT 1901  
DB 53021 TCCCTTCCAGAGTTTCCAAATTTACTATTTCTCATATCATC-----AAAGCATC 52971  
QY 1902 TCAATGCAAGTTATAGCTTATGATATGATTTCTTCAATATAAGGCTTGAAGTTGA 1961  
DB 52970 ACTGTGAGCATATAGTGTCAACAACGTATTTCTTAATATAACAAGAAATGAAGTCTAA 52911  
QY 1962 ATTAATCTCTGATTCATTTTCTGCAAAATTAATGATGTTG 2000  
DB 52910 ATGATTCCTGATTCATGAGGCTGCAAGATGATGCTGTG 52872

RESULT 15  
US-09-949-002-805/c  
Sequence 805, Application US/09949002  
Patent No. 690016  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION  
FILE REFERENCE: CL000790  
CURRENT APPLICATION NUMBER: US/09/949,002  
CURRENT FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: 60/231,401  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 10823  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 805  
LENGTH: 135687  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)-(135687)  
OTHER INFORMATION: n = A, T, C or G  
US-09-949-002-805

Query Match 22.7%; Score 453.4; DB 3; Length 135687;  
Best Local Similarity 74.7%; Pred. No. 1.1e-83;  
Matches 728; Conservative 0; Mismatches 206; Indels 41; Gaps 11;

QY 1051 ATCACAATTAAGTGTACACAGTCTTTGGCTTCCAGTGCATATTAAGTTTGTCT 1110  
DB 49232 ATGGCAATTAAGTGTGTACACAGATATTTGGTTCCAGGAGATATAAAGTTATTTT 49173  
QY 1111 TATACACACGTAGCTGTGTAAGTGTGCAATATGTTATGTTAAAAA--ACACAT 1166  
DB 49172 TAGACTATACGTAAATTTCTTAAAGTGTGCAATATGTTATGTTAAAAAATGTTAT 49113  
QY 1167 ACCCTTAATTTTAAATGCTTTTATTAATAAATGTTAAACATCAATTTGAGATTCAGTG 1226  
DB 49112 ACCATATTAATAAATAATCTTATTTGCTGAATAATGTTAAACATCAATTTGAGT 49053  
QY 1227 AGTTGATCTTTTGTGCTGTGAAGTCT---TTTCTTATGATGATGATCGGGGT 1282  
DB 49052 ACTATAAATCTTTTGTGCTGTGAAGTCTTAACTTTGATGATGATGATCGGATGAT 48993  
QY 1283 CAGG-----TGCGAAGCTTAGGGGTGCTGTGCAAGTTCT-----TAAACAAAG 1329  
DB 48992 CAGGGTGTGTTGCTGACGAGCGGGCTGTTGTGAATTTCTTAAGCATTAAGACAA 48933

OY	1330	TTGAAGATGCAATATCAAGTGAAGCTCTTCCTTTCATGAAAGATTTCTCTAGAGTGTAT	1389
Db	48932	TGAAGTCCCAACCAAGATTGACCTCTTCTTAATGATAGATTCTCTGAGCATGTGAT	48873
OY	1390	GCTTTTGAATGCAATTTTATGCAACAGTAGAATCTTTTGAAATTGGA--TCAATCTCTC	1448
Db	48872	GCTTTTGAATGCAATTTTATGCAACCAAGTAGAATCTTTCAAAATTGAGATCAATCTTTTC	48813
OY	1449	AAACCCGCTCTGCTTTAAACAACCTAGTTAATAATAATTGATGATTCATTTGTTGAT	1508
Db	48812	AAAGCGTCCACTGCTTTATTCACATGAATTTATGATATATTCATTAATCTTTGCTGCT--T	48754
OY	1509	TTCAACAATTTTCAACAGTGTCTTCACACGAGAGTAGATTTCCATCTATTTCTGAGATGGA	1568
Db	48753	TTCAACAATATTTCAACGATATTTACAGCAGGATATGATTCATCTCAAGAAACCACT---	48697
OY	1569	ATCTTTGTCTATCCATAGAAGAAATTCCTCATCTGTTCAGTTTATCATGATTTGA	1628
Db	48696	-TCTTTGTCTATCTATAGAAGCAACAGTTCATCTGTCCAAAGTTTATCATAGATTGG	48638
OY	1629	GCAATPACAGTCATGCTCTTCAGGCGCTCACTTCACTTTAAATTCAGTTCCTTGCTGCTTC	1688
Db	48637	GCAATTCAGTTACATCTTCTGG-----CTATTTTCTAATTTCAATTTCTCTGGATTTTC	48583
OY	1689	TACCAATCTGTGTTCTCTTCTCCATTTGAAGTCTTGAACTCTCAAGTCATTCATGAG	1748
Db	48582	CACAAAGATCTGCAAGTTTACTTCCAAACAGTAGTGTCTAAACCCCTTCAAATCATCATGAA	48523
OY	1749	GGTGGAGATGACCTTCTCCAAATTCCTGTAAATATTTAATTTTGA-----CCTCCCATGA	1805
Db	48522	GCGTGGATCAACTTCTTCCAAATTCCTGTTCATGTTGATTAATTTGACCTCTCCCATGA	48463
OY	1806	ATCATGATGTCTTTTAATGACACCTGGAATGTGAATCCCTTTCCAAAAGTTTCAATTT	1865
Db	48462	ATCACAATATGTTCTTAATGGAATGTGAATGTTGAGTCCCTTTCACAGGGTTTTCAATTT	48403
OY	1866	ACTTAGTCCGATGCCATTCACAGAGATTCACCTTCAATGCGCAATTTAGCCTTAATG	1925
Db	48402	ACTTTTCCAGAGCA-----CAGAGGAATCTCTAATTATGCAACTATGACCTTACGA	48348
OY	1926	AATGTATTTCTTCATATAAGGCTTGAAGTTGAATTAATCTCTGATATC--ATTTTCG	1984
Db	48347	AATGATATTTCTTAAACATATAGACTTGAAGCAGAAATCATGCTCTTGAATCCAAATGGCTG	48288
OY	1985	CAAAATGATGCTGT	1999
Db	48287	CAGATGATGTTGT	48273

RESULT 16  
 US-09-949-016-15698  
 Sequence 15698, Application US/09949016  
 Patent No. 6812339  
 GENERAL INFORMATION:  
 APPLICANT: VENTER, J. Craig et al.  
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 FILE REFERENCE: C1001307  
 CURRENT APPLICATION NUMBER: US/09/949,016  
 CURRENT FILING DATE: 2000-04-14  
 PRIOR APPLICATION NUMBER: 60/241,755  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/237,768  
 PRIOR FILING DATE: 2000-10-03  
 PRIOR APPLICATION NUMBER: 60/231,498  
 PRIOR FILING DATE: 2000-09-08  
 NUMBER OF SEQ ID NOS: 207012  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 15698  
 LENGTH: 145812  
 TYPE: DNA  
 ORGANISM: Human  
 FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: (1) .. (145812)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15698

```

Query Match	22.7%;	Score 453;	DB 3;	Length 145812;
Best Local Similarity	73.1%;	Pred. NO. 1.3e-83;		
Matches 764;	Conservative	0;	Mismatches 225;	Indels 56; Gaps 12;

QY	973	ATAGTACAGGCATACCTTGGAGTACCTGGGGTTTGCTTCATACACACACATATATACA	1032
Db	43948	AAATATCTGACACACCTCGGAGATATATGGGGTTTGCTTCAGACATATGCAAGTGAAGCA	44007
QY	1033	AATATGCAAGAAAGTGGATATACAAATPAAAGTGAATCAACAACTCTTTGGCTTCCAGT	1092
Db	44008	AAATATG-----CAATPAAAGCAAAATCAATGAAGTTTGGTTTCCAGT	44052
QY	1093	GCATATPAAAGTTTGGCTTATATCAACAGTGAAGTCTGTTAAGTGCAATAGCTATAGT	1152
Db	44053	GCATATPAAAGTTATGTTTAAATATATCTGTAAGTCTGTTAAGTATCAAAATATATAGT	44112
QY	1153	CTAAAAAACAATACCTTATATTTTAAATAGCTTTATTACTAAAAATGCTAAACATCAT	1212
Db	44113	CGAAAAAAGTACATCTTATATTTAAATATCTTTATGTCAAAAAGCACTCATATCAT	44172
QY	1213	TTGAGCATTCAGTGAAGTGTATCTTTTGTCTGGTGAAGGTCTTTCTTATG-----	1266
Db	44173	CTGAGCCTTCAGTGAAGTCAATATCTTTTGTATGGAAGTCTTGCTTGAATGTTGATG	44232
QY	1267	----ATGACGTATCGGGGGTCAAGTCTGAAGCTTAAAGGTGTGGGTGCTGGCATTTCTTAA	1322
Db	44233	GCTACGTACGTACCAAGGGTGGTGGCTGTAAGGTGGGTGCTATGGCAATATCTTAA	44292
QY	1323	ACA--ACAGTGAAGATTGCAATATGACTGATGCTCTCTTATATGAATTTCTCTCA	1380
Db	44293	ATTAAGCAATGAAGGTGTGCTGATCAAGTGAAGCTCTTGTGTGAATATGCTCTGTA	44352
QY	1381	GTGTGTGATGCTTTTGTATAGCATTTTATATGACAGTAGAAGCTCTTGAATTTGA--T	1438
Db	44353	GTATGCAATGCTATTTTGTATAGCATTTTACCCGAGTAGAAGCTTTTGAATTTAGGAAGT	44412
QY	1439	CAATCTCTCAAAACCTGCTCTGCTTTTAAACAACCTAAGTAAATATATATTTCTGAATCCA	1498
Db	44413	TGATTCACATCAAAACCTGCACTGCTT--TGATTAAGTTGATGATATATTAATCTCT	44469
QY	1499	TTGTGTGCAATTTCAACAATTTTTCACAGTGTCTTCAACAGAGTAGATTCACATCTATTC	1558
Db	44470	TTGTGTGCAATTTCAACAATTTTTCACAGATCTTCAACAGAAATGAATTCATCTCAAGAA	44529
QY	1559	CTGAGATGGAATTTTGTCTCATTCATTAAGAAATAATCTCATCTGTTCAACTTTATATCA	1618
Db	44530	ACCACTT-----TCTTTGCTCATTCATTAAGAAAGCAATCTTATATGTCCAAAGTTTATCG	44585
QY	1619	TGAGATGCAACAAATACAGTCAATGCTTCAAGGCTCATCTTCACTTTAATTCAGTTCTC	1678
Db	44586	TGAGATTAACGCAATTCAGTCAAGTCTTCAAG-----CTCCACTTCTAATTTAATGTTCTC	44640
QY	1679	TTGCTGTTTCTACACATCTGTGTTCTTCTCTCAATGAAAGCTTTGAACCTCTCAAGT	1738
Db	44641	TTGTATCTTCTACACATCTGTGAAGTGAATCTTCTTCACTGAAG--ATTAATATCTCTCAAGAA	44699
QY	1739	CATCATGAGGGTTGGAATCGACTTCTTCAAAATCTGTTAATATTTATATTTGACCT	1798
Db	44700	TACCATGAGAGCGCAATATCAAGCTTCTTCCAAATCTCTGTTATATTTGACATTTTGACCT	44759
QY	1799	CC---CATGAATCATGAATGTTCTTATATGCACTGGAATGTTGAATCTTTTCCAAAAG	1855
Db	44760	CCCTCATGAATCAAAATGTTTTCATGAGCTCTGAAGTGGTGAACCTTTTCCATAAGT	44819
QY	1856	TTTTTCAATTAATTAAGTCCAGATTCATCATCAGAGATTCACCTTTCAATGCCAGTTAT	1915
Db	44820	TTTTTATTTATCTTGCTTAAGATCTGT-----GAAGAGAAACACT---GTGACGCAAT	44870





Db 28076 AAGTGAATATGCAATTAAGACAGTCAAGATAATTTGGTTTCCAGTGCATATATAA 28017  
 Qy 1103 GTTTGCTTATACACTAGTCTGTTAAGTGCATAGTGTATGCTTAATAAAC 1162  
 Db 28016 GCTACATTTACATTACACATGCTATTAATAGGGCAATAGATATGTTAAAAA--- 27960  
 Qy 1163 ACATACCTTAATTTAAAAATGCTTTATTAATAAAAAATGTAACAATCAATTTGAGCATTC 1222  
 Db 27959 ATACACTTGTATTAATAAATCTTTTATTAATAAAAAATCAATCACTGAGCTTG 27900  
 Qy 1223 AGTGAATGTATCTTTTGTGCTGTAAGTGTCTTTTCTTATGTATGATGATCGGGGT 1282  
 Db 27899 AGCAATGTGTAATCTTTTGTGCTGCTGCTCATTTGATGATGCTGCTAGT--GGAG 27842  
 Qy 1283 CAGGTGCTAAGCTTGGGTGCTGAGAGTTCTTTAA-----CAACAGAGATTT 1337  
 Db 27841 TCTTTGCTAAGTGGGGGTGCTGATAGCAATTTCTTAATAAGGCAACAAAGAGCTT 27782  
 Qy 1338 GCAATATCAGTGAATCTCTCTCTTCAAGAAATTTCTCTAGTGTGATGATCTTTTG 1397  
 Db 27781 CCCACATCACTGACTCTT-CTTTGATGATGATTTGTTGATGATGATGATGATGATG 27723  
 Qy 1398 ATAGCATTTTATGACAGTAGAACTTTCTTGAATAATGGA-TCATCTCTCAACCTTG 1456  
 Db 27722 ATAGCATTTTATGACAGTAGAACTTTCTTGAATAATGGA-TCATCTCTCAACCTTG 27663  
 Qy 1457 CTCTGCTTAACTAAGTAT 1516  
 Db 27662 AGCTGCTTATCACTAAGTATATATATATATATATATATATATATATATATATATAT 27606  
 Qy 1517 TTTTCAAGTGTCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1576  
 Db 27605 TCTTCAAGATTTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 27547  
 Qy 1577 TCATCATTAAGAAATTCCTATCTGTTCAAGTTTATCATGATGATGATGATGATGATGAT 1636  
 Db 27546 TCTTTAAGAAAAAGAACACCTCATCATCTTCAAGTTTATCATGATGATGATGATGAT 27487  
 Qy 1637 GTCAATGCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 1696  
 Db 27486 GTCAATGCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 27432  
 Qy 1697 CTGATGCTTCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 1756  
 Db 27431 CTGATGCTTCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 27372  
 Qy 1757 TCGACTTCTTCAAAATTCCTGTTAATATATATATATATATATATATATATATATATAT 1813  
 Db 27371 TCAACTTCTTCAAAATTCCTGTTAATATATATATATATATATATATATATATATAT 27312  
 Qy 1814 TGTTCCTTATGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1873  
 Db 27311 TGTTCCTTATGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 27252  
 Qy 1874 CAGATCCATCCATCCAGAGATCCATTCATGATGATGATGATGATGATGATGATGATGAT 1933  
 Db 27251 CAGATCCATCCATCCAGAGATCCATTCATGATGATGATGATGATGATGATGATGATGAT 27201  
 Qy 1934 TCTTCAATATATAGCTTGAAGTGAATATATATATATATATATATATATATATATATAT 1993  
 Db 27200 TCTTCAAT 27143  
 Qy 1994 TGTTCGTG 2000  
 Db 27142 TGTTCGTG 27136

RESULT 19  
 US-09-949-016-13249/C  
 ; Sequence 13249, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH LOCAL INFECTION WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CLO01307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13249  
 ; LENGTH: 264304  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-016-13249

Query Match 22.2%; Score 444.2; DB 3; Length 264304;  
 Best Local Similarity 74.9%; Pred. No. 9,76-82;  
 Matches 724; Conservative 0; Mismatches 208; Indels 35; Gaps 12;

Qy 1043 AAGTGAATATGCAATTAAGACAGTCAAGATAATTTGGTTTCCAGTGCATATATAA 1102  
 Db 28076 AAGTGAATATGCAATTAAGACAGTCAAGATAATTTGGTTTCCAGTGCATATATAA 28017  
 Qy 1103 GTTTGCTTATACACTAGTCTGTTAAGTGCATAGTGTATGCTTAATAAAC 1162  
 Db 28016 GCTACATTTACATTACACATGCTATTAATAGGGCAATAGATATGTTAAAAA--- 27960  
 Qy 1163 ACATACCTTAATTTAAAAATGCTTTATTAATAAAAAATGTAACAATCAATTTGAGCATTC 1222  
 Db 27959 ATACACTTGTATTAATAAATCTTTTATTAATAAAAAATCAATCACTGAGCTTG 27900  
 Qy 1223 AGTGAATGTATCTTTTGTGCTGTAAGTGTCTTTTCTTATGTATGATGATCGGGGT 1282  
 Db 27899 AGCAATGTGTAATCTTTTGTGCTGCTGCTCATTTGATGATGATGATGATGATGATG 27842  
 Qy 1283 CAGGTGCTAAGCTTGGGTGCTGAGAGTTCTTTAA-----CAACAGTGAATTT 1337  
 Db 27841 TCTTTGCTAAGTGGGGGTGCTGATAGCAATTTCTTAATAAGGCAACATGAGCTT 27782  
 Qy 1338 GCAATATCAGTGAATCTCTCTCTTCAAGAAATTTCTCTAGTGTGATGATGATGATGAT 1397  
 Db 27781 CCCACATCACTGACTCTT-CTTTGATGATGATTTGTTGATGATGATGATGATGATGAT 27723  
 Qy 1398 ATAGCATTTTATGACAGTAGAACTTTTGAATAATGGA-TCATCTCTCAACCTTG 1456  
 Db 27722 ATAGCATTTTATGACAGTAGAACTTTTGAATAATGGA-TCATCTCTCAACCTTG 27663  
 Qy 1457 CTGATGCTTCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 1516  
 Db 27662 ATGCTGCTTCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 27606  
 Qy 1517 TTTTCAAGTGTCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 1576  
 Db 27605 TCTTCAAGCAATTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 27547  
 Qy 1577 TCATCATTAAGAAATTCCTATCTGTTCAAGTTTATCATGATGATGATGATGATGATGAT 1636  
 Db 27546 TCTTTAAGAAAAAGAACACCTCATCATCTTCAAGTTTATCATGATGATGATGATGATGAT 27487  
 Qy 1637 GTCAATGCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 1696  
 Db 27486 GTCAATGCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 27432  
 Qy 1697 CTGATGCTTCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 1756  
 Db 27431 CTGATGCTTCTTCAAGGCTCATCTTCAATTTTCAAGTTTCTGCTGTTTCTTCAACAT 27372  
 Qy 1757 TCGACTTCTTCAAAATTCCTGTTAATATATATATATATATATATATATATATATATAT 1813



Db 27371 TCACTTCTTCCAAACCTTCTCTATGTGTGATGTTTGACATCTCCCATGATCATGAA 27312  
Qy 1814 TGTCTTAAATGACACCTGGATATGGTGAATCTCTTCCAAAGGTTTCAATTACTTATGTC 1873  
Db 27311 TGTCTTAAATGATGATCTCAAAATGATATCCCTTCCAAAGGTTTCAACTCACTTATGTC 27352  
Qy 1874 CAGATTCATCATCGACAGATATCCACTTTCATATGCGACGTTATAGCTTATGAAATGTAAT 1933  
Db 27251 CAGATTC-----CGAGGATCACT-----ATGCGAGTATAGCTTATGAAATGTAATG 27201  
Qy 1934 TCTTCAATATATAGGCTTGAAGTTGAATTAATCTCTTGATTCATTTCTGCAAAATGTA 1993  
Db 27200 TCTTAAATATAT--GACTTGAATTTTGAATTTCTCTGATCCAAAGGACGTGTAAGGA 27143  
Qy 1994 TGTGTG 2000  
Db 27142 TGTGTG 27136

RESULT 20  
US-09-949-016-14125/C  
; Sequence 14125, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14125  
; LENGTH: 177797  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(177797)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14125

Query Match 22.1%; Score 443; DB 3; Length 177797;  
Best Local Similarity 73.0%; Pred. No. 1,6e-81;  
Matches 725; Conservative 0; Mismatches 215; Indels 53; Gaps 10;

Qy 1043 AAGTGATATCAATATTAAGTGTGACACAAAGTCTTTGGCTTCCAGTCAATATAAA 1102  
Db 11287 AAGTATATATCAATATTAAGTGTGACACAAAGTCTTTGGCTTCCAGTCAATATAAA 11228  
Qy 1103 GTTTTGCTTAACTAATCTGTAGTCTGTATTAAGTGTGCAATAGTATGCTTAAATAAA 1162  
Db 11227 GTTATGTTTAACTAATCTGTAGTCTGTATTAAGTGTGCAATAGTATGCTTAAATAA 11170  
Qy 1163 ACATACCTTAATTTTAAATGCTTATTTAC-TAAATAATGCTAATCATTTGAGCAT 1221  
Db 11169 ACATACCTTAATTTTAAATGCTTATTTGCTTAAATAATGCTAATCATTTGAGCAT 11110  
Qy 1222 CAGTGTGTATCTTTTGTGTGTGAGGAGTCTTTTCTTATGTATGA-----1270  
Db 11109 CTGTGTGTATATCTTTTGTGTGTGAGGAGTCTTTGCTTATGTATGTATGTATGAAC 11050  
Qy 1271 CTGT 1325  
Db 11049 CAATCAGATGT 10990  
Qy 1326 ACAGTGAATGT 1385

Db 10989 ACAGTGAATGT 10930  
Qy 1386 TGATGCTTTTGTATGATTTTATGACAGATGAATCTTTTGAATAATGGATCAATCT 1445  
Db 10929 TGATGCTTTTGTATGATTTTATGACAGATGAATCTTTTGAATAATGGATCAATCT 10871  
Qy 1446 CTCAAACTGTCTGTCTTAAACAATGTTAATATATATTTGTATGATTCATTTGTGT 1505  
Db 10870 CTTAAGTCTGTCTGTCTTAAACAATGTTAATATATATTTATGATTCATTTGTGTGT 10811  
Qy 1506 CATTTCAACAATTTTCAAGTGTCTTCAACGAGATGATTTCCATTCATTTCTGTGAT 1565  
Db 10810 CATTTTCAACAATTTTCAAGTGTCTTCAACGAGATGATTTCCATTCATTTCTGTGAT 10752  
Qy 1566 GGAATCTTTGTCTGTCTTAAACAATGTTAATATATATTTGTATGATTCATTTGTGT 1625  
Db 10751 -----TCTTTGTCTGTCTTAAACAATGTTAATATATTTATGATTCATTTGTGT 10696  
Qy 1626 GCAGCAATATAGTATGTCTTCAAGGCTTCACTTCACTTTAATTCAGTCTTCTGTGT 1685  
Db 10695 GCAGCAATATAGTATGTCTTCAAGGCTTCACTTCACTTTAATTCAGTCTTCTGTGT 10641  
Qy 1686 TTCTACCAATCTGTGTGTCTTCTCTCATTTGAAGTCTTGAACCTTCCAGTATCAT 1745  
Db 10640 TTCCACCAATCTGTGTGTCTTCTCTCATTTGAAGTCTTGAACCTTCCAGTATCAT 10581  
Qy 1746 GAGGTTGGATTCGACTTCCAAATCTCTGTTAATATTTATTTTGA-----1795  
Db 10580 AAGGTTGGATTCGACTTCCAAATCTCTGTTAATATTTATTTTGA-----1795  
Qy 1796 -----CTCCCATGATCATGATGTTCTTAAATGACCTGATGATGATCTTT 1847  
Db 10520 TTTGACCTCTCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10461  
Qy 1848 CCAAAAGTTTCAATTTTACTTATGTCGATTCATTCATTCAGAGATTCATTCATGAT 1907  
Db 10460 CCAAAAGTTTCAATTTTACTTATGTCGATTCATTCATTCAGAGATTCATTCATGAT 10406  
Qy 1908 CCAGTTATAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1967  
Db 10405 CCAGTTATAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10346  
Qy 1968 CCTGATCATTTTCTGCAAAATATGATGTTGTG 2000  
Db 10345 CCTGATCATTTTCTGCAAAATATGATGATGATGATGATGATGATGATGATGATGAT 10313

RESULT 21  
US-09-949-016-11842/C  
; Sequence 11842, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11842  
; LENGTH: 227979  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature

LOCATION: (1)....(227979)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-11842

Query Match 22.1%; Score 443; DB 3; Length 227979;  
Best Local Similarity 73.0%; Pred. No. 1.7e-81;  
Matches 725; Conservative 0; Mismatches 215; Indels 53; Gaps 10;

1043 AGTGGATATTCACATTAAGTACACAGTCTTTGGTCCCAATGATATAA 1102  
13469 AAGTATTAATACAAATAAGGACCTACATGAACTGTGAGTTCCCAATGATAA 13410  
1103 GTTTGCTTATACACTAGTGTGTTAAGTGTGCAATAGTATGCTAAAAA 1162  
13409 GTCAATGTTTACACTAGTGTGCTATTAAGTTGCAATGCGTATGCTT- 13352  
1163 ACATACCTTAATTTTAAAAAGCTTTATAC-TAAAAATGCTAACATCTTGA 1221  
13351 ACATACCTTAATTTTAAAAAGCTTTATGCTTTAAAAATGCTAATCATG 13292  
1222 CAGTGAATGTTATCTTTTGTGTGGAAGCTTTTCTTATGATGA- 1270  
13291 CTGTGATCATATACCTTTTGTGTGGAAGGCTTTGCTGATGTTGATGA 13232  
1271 CTGATCGGGGTACAGTGTGGAAGCTTGAAGTGTGAGCTTTCTTAA- 1325  
13231 CAATAGAGATGGGGTCTGTAGGGTGGGCTGTGCAATTTCTTAAATAAG 13172  
1326 ACAGTGAAGATGCAATATCACTTGAATCTTCTTCAATGAAGATTTCT 1385  
13171 ACAGTGAAGTGTGAGCTTACCTGACCTTCTTAAATATTTCTGTGTA 13112  
1386 TGATCTTTTATAGCATTTTATGACAGTGAAGTCTTTGAAATGATCAT 1445  
13111 TGATCTTTTATAGCATTTTATGACAGTGAAGTCTTTGAAAT- 13053  
1446 CTCAACCTGCTGCTTAACTTAACTTAACTTAACTTAACTTAACTT 1505  
13052 CTTAAGTCTGCTTAACTTAACTTAACTTAACTTAACTTAACTTAA 12993  
1506 CATTTCAATTTTCAAGTGTCTTCAACAGAGTGAATTCATCTCAATCT 1565  
12992 CATTTCAACATTTTCAAGTGTCTTCAACAGAGTGAATTCATCTCA 12994  
1566 GGAATCTTTGCTCATATGAGAAATCTTCACTGTGTTAAGTTTATG 1625  
12993 -TCCTTGTCTCATATGAGAAATCTTCACTGTGTTAAGTTTATG 12878  
1626 GCAGCAATACAGTGTCTTCAAGTGTCTTCACTGTGTTAAGTTTATG 1685  
12877 GCAGCAATACAGTGTCTTCAAGTGTCTTCACTGTGTTAAGTTTATG 12823  
1686 TTTTCAACATCTGAGTGTCTTCAAGTGTCTTCACTGTGTTAAGTT 1745  
12822 TTTTCAACATCTGAGTGTCTTCAAGTGTCTTCACTGTGTTAAGTT 12763  
1746 GAGGTTGGAATGCACTTCTTCAATCTTCTTAAATTTATTTTGA- 1795  
12762 AAGGTTGGAATGCACTTCTTCAATCTTCTTAAATTTATTTTGA 12703  
1796 -CTTCCCAATGATGCACTTCTTCAATCTTCTTAAATTTATTTTGA 1847  
12702 TTTGACCTTCTTCAATGATGCACTTCTTCAATCTTCTTAAATTT 12643  
1848 CCATAAGTGTCTTCAATGATGCACTTCTTCAATCTTCTTAAATTT 1907  
12642 CCATAAGTGTCTTCAATGATGCACTTCTTCAATCTTCTTAAATTT 12588  
1908 CCAGTTATAGCTTATGAGTATTTCTTCAATGATGCACTTCTTAA 1967  
12587 GCAGTTATAGCTTATGAGTATTTCTTCAATGATGCACTTCTTAA 12528  
1968 CTTTATCATCATTTCTGCAAAATAGATGTTG 2000

Db 12527 CCTTATCATGAGGCTACAGAAATGATGTTG 12495

RESULT 22  
US-09-497-855A-38  
Sequence 38, Application US/09497855A  
Patent No. 6605432  
GENERAL INFORMATION:  
APPLICANT: Huang, Tim  
TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION  
FILE REFERENCE: UMO1523  
CURRENT APPLICATION NUMBER: US/09/497,855A  
CURRENT FILING DATE: 2000-02-04  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: 60/118,760  
PRIOR FILING DATE: 1999-02-05  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 38  
LENGTH: 128779  
TYPE: DNA  
ORGANISM: Homo sapiens;  
US-09-497-855A-38

Query Match 22.0%; Score 439.8; DB 3; Length 128779;  
Best Local Similarity 72.7%; Pred. No. 6.7e-81;  
Matches 747; Conservative 0; Mismatches 232; Indels 48; Gaps 12;

977 TACAGCATACCTTGTAGTACTGTGGTGTCCATACACCAATATATACAA 1036  
42749 TACAGCATACCTTGTAGTACTGTGGTGTCCATACACCAATATATAC 42808  
1037 TGCAGGAATGATATACCAATTAAGTGTGACCAAGTCTTTGGTCCAG 1096  
42809 TTG-----CAATTAAGTGTGACCAATTTTGGCTTCCATGAT 42853  
1097 ATAAAGTTTGTCTTATACATCTGATGTCTTAAAGTGTGCAATGTTAT 1156  
42854 ATAAAGTTTGTCTTATACATCTGATGTCTTAAAGTGTGCAATGTT 42911  
1157 AAAACATATCTTAAATTTAAATGCTTTATCTTAAATGCTTAAATGCT 1216  
42912 AAATCATATATCTTAAATTTAAATGCTTTATCTTAAATGCTTAAAT 42970  
1217 GCATTCAGTGTGATTAATCTTTTGTGTGGAAGTCTTTCTTATGATGA 1275  
42971 GAGTCATATATCTTGAAGTGTGCAAGTCTTAAATGCTTGAATGCTGA 43030  
1276 CGGGGTGAGTGTGCAAGTCTTGAAGTGTGCAAGTCTTGAAGTCTTGA 1330  
43031 AGGGGTGATTAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 43090  
1331 GAAATGCAATATGATGTTGCTTCTTCAATGAAGTGTGCTTCAATGAT 1390  
43091 GAAATGCAATATGATGTTGCTTCTTCAATGAAGTGTGCTTCAATGAT 43150  
1391 CTTTGTATGATGATTTATGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 1450  
43151 CTTTGTATGATGATTTATGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 43205  
1451 ACCGTGCTGCTTAAACCTTAAGTGTGCAAGTGTGCAAGTGTGCAAGT 1510  
43206 ACTGTGCAATGCTTAAACCTTAAGTGTGCAAGTGTGCAAGTGTGCAAG 43265  
1511 CAACATTTTCAAGTGTGCTTCAACAGGATGATGATGATGATGATGATG 1570  
43266 CAACA-TTTCACACATCTTCCAGGATGATGATGATGATGATGATG 43320  
1571 CTTTGTATGATGATTTATGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 1630  
43321 CTTTGTATGATGATTTATGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 43380



```

; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13969
; LENGTH: 301828
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc.feature
; LOCATION: (1)...(301828)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13969

Query Match      21.9%; Score 437.4; DB 3; Length 301828;
Best Local Similarity 69.9%; Pred. No. 2.5e-80;
Matches 797; Conservative 0; Mismatches 261; Indels 83; Gaps 12;

QY      875 TTTTATGCTATGCTATATATACATTAGTTGTTGTTGTTGTTTTCCTTATGCTG 934
DB      224737 TTTTCTGTTCTGTTGTTGACAGGCGATTAGTTCCTTCTGAGATTAATGATCCAA 224796

QY      935 TTGGCTGGGGTCAGCAAACTTTTCTGTAAAGGGCTAGATAGTACAGCATCTTGGAG 994
DB      224797 GTGGCATTTCTTGCTCTGCGAAATCCCAAAACAAGATCTTTGGACGACACTCAGAG 224856

QY      995 ATACTGTGGGTTGGTTCCATACACACACATATATCAAAATATGACAGAAGTGATATCA 1054
DB      224857 ATATGTCAGAGCTCAGTTCCAGACACCAACA-----AAGTAATATCG 224901

QY      1055 CAATAAGTGACACACAAGCTTTTGGCTTCCAGTGATATAAGTTTGGCTATA 1114
DB      224902 CAATAAGGACAGTAACACACATTTTCTT-----AGTGATATAAGTTAGTTTACA 224957

QY      1115 TCACTGTAGTCTGTAAAGTGACATAGTGTATGCTAAAAAAACACATACCTTAAT 1174
DB      224958 CTCTTCTAGTCTATTAATAGTGT-----ACATACCTTCAAT 224994

QY      1175 TTTAAATGCTTTATTAATAAATATGCTAACATATTAAGTATGAGTGTGTA 1234
DB      224995 TAAATAA---TTTACTCTTAAATAATGCTAGTATCATCTGATCATGATGATTTTAA 225051

QY      1235 TCTTTTGGTGGTGAAGGCTTTTCTTATGATGA-----CTGATGGGGGTCA 1284
DB      225052 TCTTCTTCTGAGAGGGCTTCTGCTCAGTGTGCTGCTGCTATCGGGGGTg 225111

QY      1285 GGTGCTGAAGCTTAAGGTGGCTGTGSCAGTTTCTTAA-----ACAACAGTGAAGTTGC 1339
DB      225112 GTAGCTGAAGATTGTGTGGTGGCGCAATTTCTTAAATTAAGACACAAATGAAGTTGC 225171

QY      1340 AATATCAGTGAATCTTCTCTTTCATGAAGATTTCTCTAGTGTGATGCTTTTGAAT 1399
DB      225172 TGCATGCACTGACTCTTCTCTTGAACAAAGATTTCTCTGATGATGTGATGCTTGAAT 225221

QY      1400 AGCATTTTATGACAGTGAATCTTCTTGAATTTGA--TCAATCTCTCAAACTGCT 1458
DB      225232 ACCATTTTACCAAGATGAATCTTCTTCAAAATTTGAGTCAATCTTTCATATCTGCT 225291

QY      1459 CTGCTTTAACAACCTAAGTTAATATATATATTCGANTTCATGTTGTCAATTCACAAAT 1518
DB      225292 GCTGCTTCACTACATAGATTATATATCTTCTTAAATCTTGTGTGATTTCAATTCACAAAG 225351

QY      1519 TTCACAGTGTCTTACACAGAGTGAATTCATCTCATTTCTGAGATGGAATCTTTGCTC 1578
DB      225352 CTCACAGATCTTCACTAGAGAGTCTTCATCTCAAGAGACACTT-----TCTTTGCTA 225407

QY      1579 ATCCATTAAGAAATTTCTCATCTGTCTCAAGTTTATATAGATTTGACGCAATTCAGT 1638
DB      225408 ACCCGTAAGAGCACTCTCATATGTTCAAAGTTTATATCATGATGTTGACGCAATTCAGT 225467

QY      1639 CATGTCTTACGGGCTCACTTCACTTTAATTCAGTTCTCTGCTGTTTCTACACATCT 1698
DB      1639 CATGTCTTACGGGCTCACTTCACTTTAATTCAGTTCTCTGCTGTTTCTACACATCT 1698
```

```

DB      225468 CACAGTTGAG-----CTCCACTTCTAATTCATTTCTTACTTTTCCACACATCT 225522
QY      1699 GTGGTTCCTCTCCATTTGAAGTCTGAACCTCTCCAGTCAATGAGGGTTGAATC 1758
DB      225523 GCTGTACTCTCTCCACTGAAGCTTAAAGCCCTCAAGTCAATCAGTGGGTTGAAT 225582
QY      1759 GACTTCTTCAAAATCTCTGTATATATTTATTTTGAACCTCCATGAATCATGATGTT 1818
DB      225583 CACTTCTTCAAACTCTCTGTATATTTTCACT---CTCCATGAATCAGCAATGTT 225639
QY      1819 TTAATGGACCTGGAATGTGAATCTTTCAAAAGTTTCAATTTACTTACTGACAT 1878
DB      225640 TTAATGACATTAAGATGCGAATCTTTTCAAGATTTTCAATTTATTTGCCAGAT 225699
QY      1879 CCATCATCCAGAGATTCACATTTCAATGCCAGTTATACCTTATGGAATGATTTCTTC 1938
DB      225700 TCATC-----AAAGAAATCAGATGAGTGTAG--CTTATTAATTAATTTCTTA 225749
QY      1939 AATAATAGGCTTGAAGTTGAATTAATCTCTTGATCCATTTTCTCAAAATGATGTTG 1998
DB      225750 AGTAATAGCTTTGAAGATCAAAATTAATTAATTAATTAATTAATTTCTTA 225809
QY      1999 T 1999
DB      225810 T 225810

RESULT 25
US-09-949-016-16073/c
; Sequence 16073, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16073
; LENGTH: 385136
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc.feature
; LOCATION: (1)...(385136)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16073

Query Match      21.9%; Score 437.4; DB 3; Length 385136;
Best Local Similarity 57.6%; Pred. No. 2.6e-80;
Matches 1165; Conservative 0; Mismatches 746; Indels 110; Gaps 17;

QY      3 ATTTCCCGCCCTCAGCTCCCAAGTACGAGTTACAGGTCCAGCCACGCTCCAGC 62
DB      313940 ATTTCTCTCCCTCAGCTCTCAAGTACGAGTTATAGGCAATGACCAACGCGCGC 313881
QY      63 TAAATTTTGTATTTTATAGAGACGGGGTTTCAACGTTTTCAGAGCTGTATCAAACT 122
DB      313880 TAAATTTTGTATTTTATAGAGATAGGGTTTCTCATGTTGATCAGGCTGTGTAAT 313821
QY      123 CTGACCTCAAGTATCTGCTGCTCCTCAGCTCCCAAAATGCTAGATTAACAGGTGTAG 182
DB      313820 CCCAACCTCAGGTGACCACTCTGCGCTGCTCCCAAAAGTGTGAGATTAACAGGTGTAG 313761
QY      183 TCACCGCACCCAGCCCTTCTTCACTTATCACTCTTTTGTGCTATATTTGTATGAGAG 242
DB      183 TCACCGCACCCAGCCCTTCTTCACTTATCACTCTTTTGTGCTATATTTGTATGAGAG 242
```

D	b		313760	CCACGCGGCCGC-----GTGGATCTTTTTTTTTTTTTTTTTTTTTCCTTGATCTCAT	313707
O	y		243	CTTTATTATTAGGGGCACTACATTAATAATGTATGCTTATTTAGATTAATGAATGATCTGT	302
D	b		313706	TATCTTCCTATCTCTCACAGCATCAGAACAAGTAAACAGAAATGCTCAGCCAGGCCG	313647
O	y		303	CATTATGAATGCTCTGATTTCAT---TCCCTGATAGTATTCTTTTCTTAATATTTTCT	359
D	b		313646	AGTGGCTCACGCTGTAAATCCCGGCACTTTGGAGGCCAGAGGGGCAATCACTTGAGG	313587
O	y		360	GAAATGCTGCTATTAAATAGCCACTCGGCTTTTAAAATTAGATTTTATGSTAT	419
D	b		313586	TCAGAGATTGAGACCAGCTGGCCAATGTGAAAAACCCATCTCTACTAAAAATACAA	313527
O	y		420	AATATTTTCCCTTTTTTTTTTTTTTTTAAAGTTTAAAGTTATGTTTCTTATCTTAAAGT	479
D	b		313526	AAATTAGCAAGGCAATGCTGGCGGCACTAATATCCAGTTACTCGGAGGCTGAGGCAG	313467
O	y		480	GGTGTCTTATPAGCAGCATATATCTGGGCTTGATGATATTATTTAAATCTGATACTCAA	539
D	b		313466	GGAATTTGCTTTAACACCCAGAGGCAAGGTTCATATGAGCTAAGATCTGTCCATTCGACT	313407
O	y		540	CCTTTTTGTGGAGNGTTTAGGCCATTTTCATTAGAGTAAATATA-GACATGTTTTGAT	598
D	b		313406	CCAGCTGGGCGACAGACGAAAGCTGMAAAAAACAACAAAAAACAAAAAAGTAA	313347
O	y		599	TTGCTATACCATCTTTTCATTTGTTTATATATGAGGACATCTTTTCATTTGTTCTTTTC	658
D	b		313346	ATGCTCAACAAATGTTAGTGATGATGAAGAAATCCAGTAATCTAAGCATGTGTTTC	313287
O	y		659	ATCTTTGACCAATTTTCTTATAGTACGAATACITTTTGTGATTTCATTATATCTATTGGC	718
D	b		313286	ACCTATCATC-----CTGAAGCCCTGCTGTGTAAAGCCACCTGATCCT	313244
O	y		719	TTTTTAGTATACCTTTAAATTTTTTTTTTTCGTTTATGTPAGATTTTAAATATACAT	778
D	b		313243	GAATTTGGAGTCCCAAGGCTTGATCATTCACCTGTGTCACAAAGCTTCTGGAACCTT	313184
O	y		779	CTTTAATCTATCAAGATTACCTTCAATAATGATATTTTACAGCTCAAGSTAAATGTAAGA	838
D	b		313183	ACTGCTCATCTTTGTTATTATATATAGCCTTAAACAAGTACAAAGATTCGAATATCCAT	313124
O	y		839	ACCTTACAAAGATATATTTTCATTTCTGTCTCTTAATTTTATGCTATGCTATATATACA	898
D	b		313123	AGTTCCTGGACATCATTTCTCTCGAAGACGTATTTTGTGATTTATTTTAACTC	313064
O	y		899	TTAGGTTTGTGTGTTTGTTTTAACTTATGCTGTGTGGCTGGGTCACCAAAACATTTT	958
D	b		313063	TCYTAGTTATTTTCAAGCAACAAATGAATTAATTAAGTATCTTGTAAATCATTTAACATAA	313004
O	y		959	CTGTAAAGGCTAGATATGTCACAGGCATACCTTGAGATATCTGTGGTTTGTTGCATACC	1018
D	b		313003	GTATCTTTTATACATATCTATAGGCACCTTTGGAATATATGTGAGTTCAAGTTCCAGACC	312944
O	y		1019	ACCAACATATATACAAATATGCAAGAATGAGATATCACATTAAGTGAATCAACAACTT	1078
D	b		312943	ACTACATATAAGCAAAATATCCCAATAAAGTTAGCACACTTTTTT-----TTT	312896
O	y		1079	TTTGGCTTCCCAGTGCATATTAAGTTTGTCTTATACTACAGTACTCTGTAAAGTGTG	1138
D	b		312895	TCGTGTTTCTCAGTGCATATTAAGT-----TTACACTGACGTACTGTGTAAAGTGTG	312844
O	y		1139	CATATGTTATGCTTAAAAAAACACATACCTTA-----ATTTPAAANGCTTATTTACTA	1194
D	b		312840	CACACACCTTGTCTTCAAAATTAATATATATATATCATTAATTAATAATGTTTATTTGCTA	312780
O	y		1195	AAAAATGCTAACATCATTTTGACATTCAGTGAATGTAATCTTTTGTCTGTGGAAG-	1253
D	b		312780	AAAAATGCTACATCATCTGAGCTTCAAGTGAATGTAATCGTTTCTGTGTGGAAGGC	312722
O	y		1254	-TCCTTTCTTATGATGACTGATCGGGGGTTCAGG-----TGTGAAGCTTAAGGTTGGCTG	1307

```

Db      312720 TGGTCTTGATGTTGATAGTACTGCTGCTGATCAAGGGTAGTGGCCAAAGGTGAGTGCGCTG 312661
QY      1308      TGGCGAGTTCTTTAAA-----ACACAGCGAAGATATGGCAATATCAATTAATGACTCTTCCTTC 1362
Db      312660 TGGCAATTTCTTTAAATAATGACCAACAATGAATATGCGCGATCAATTA--CTTCTTTTC 312603
QY      1363      ATGAAGAATTTCTCTCTAGTGTGTGATGCTTTTGTATAGCATTTTATGCAAGTGAAC 1422
Db      312602 ACAAAAGATTTCTCTGTAGCATACATGCTGTTTGTAGCATTTAGACCAAGAGAACT 312543
QY      1423      TCTTTGAAAATTTGA--TCAATCTCTCAAAACCCTGCTCTGCTTTAACAACCTAAGTTAAT 1481
Db      312542 TCTTTAAAATATGAGATCAATCTCTCAAACTCTGCGCACTTATCAACTAAGTTAAT 312483
QY      1482      ATAAATTTCTGAATTCATTTGTTGTTCATTTTCAACAATTTTACAGTCTTACCCAGAGT 1541
Db      312482 GTCATATTTTAAAGC-----TTTTTCCCAAACT 312454
QY      1542      AGATTCACATTCATTTCCGATGAGAAATCTTGTCTCATTCATTAAGAAATATCTCTCAT 1601
Db      312453 AGATTCATCTCAAGACACACTT-----TCTTTGATCATCCGAAGAAAGCAATCTCTCAT 312398
QY      1602      CTGTTCAGTTTATCATGAGATGATGACCAATACAGTCATGTTTCAAGGCTCACTTCAAC 1661
Db      312397 CCATTCAGTTTATCATGAGACTGACAGACTGTCAGTCACATCTTCAGTCTCACACTTC-- 312340
QY      1662      TTTTAAATTCAGTTCTCTTGTGCTGTTTCTTACCAATCTGTGTGTTCTCTCTCCATGAAGT 1721
Db      312339 --TAAATATAGGTCTCTGTGCTGTTCACCAATCTGAGTTAAATTTCTCCCTGAAAT 312283
QY      1722      CTGTAACCTCTCAGATCATCCATGAGGGTGTGAATTCGACTCTTCCAAATCTCTGTAA 1781
Db      312282 CTGTAACCTCTCAAGGTCAACCCATGAGGTTGAAATCAATCTTTCCAAACCTCCGCTAA 312223
QY      1782      TATTTAATATTT--TGAACCTCCATGATCATGATGATGTTCTTATGAGCACTGAAATGTTG 1839
Db      312222 AATATGATTTTACCCCTCCCTGATGATCAAAATGTTCTTGATGGCATCTGAATATAGTG 312163
QY      1840      AATCTTTCCAAAGGTTTTCATTTACTTATGTCAGATTCATTCATCCAGAGATCCAC 1899
Db      312162 AATCTTTCTCGAAGGCTTTCATTTACTTCCCAAGATCCAT----CAGAGGATATCC 312108
QY      1900      TTTCATGCGCAGTTATAGCGCTATAGGAATGATTTCTTCATATATAGGCTTGAAGTTG 1959
Db      312107 TATCATGCGCAGCTACAGCTTACAGAAATGATTTCTGATATATATACGACTTGAAGTCA 312048
QY      1960      AATATTCTCTGATTCATTTTCTGCAAAATAGATGTTGTG 2000
Db      312047 AATATTCTCTGATTCATGAGGCTGAGAAATGATGTTGTG 312007

RESULT 26
US-09-949-016-16433/C
; Sequence 16433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL0010307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16433
; LENGTH: 96202
; TYPE: DNA

```





||||| 44317 CAACCTCCAAACCAATGCTGCTTTAATGA-CAAAATTTTGTAGTATCCATAATCTTT 44375  
||||| 1500 TGTGTCAATTTCAACAATTTTTCACAGTGTCTTCACACAGAGATAGATTCATCTCATTTCC 1559  
||||| 44376 TGTGTATTTTAAATATATATATACAGCATCTTCACAGAGATTTGATC-----TTCA 44426  
||| 1560 TGAGATGGAATCTTGTCTCATCATAGAGAAATTCCTCATCTGTCCAAAGTTTATCAT 1619  
||| 44427 GAAACCAATCTTGTCTGTCTCATTTGAGAACCTCTTATTCATTCATCAAGTTTGATTA 44486  
||| 1620 GAGATTCAGCAATACAGTATGTCTTCAGGCTCACTTCATCTTTAATTCAGTTCTCT 1679  
||| 44487 GAGATGGACAGGTCAGTCAATTTTCAAG-----TTCACCTTCAATTTCAAGTTCTCT 44541  
||| 1680 TGTCTTTCTCAACATCTGTGTTCTCTCTCATTTGAAGTCTTGAACCTCTCCAAAGTC 1739  
||| 44542 TGTCTATTTCCATGTCATCTGCGAGTCTTCTCTCCACTGAAGTCTTGAACCTCTCAAAAGTC 44601  
||| 1740 ATCCATGAGGAGTTGGAATGACCTTCTCCAAATTCCTGTAAATTAATTAATTTTG---AC 1796  
||| 44602 ATCCATGAGAGTTGGAATGACCTTCTCCAAATTCCTGTAAATTAATTAATTTTGACTT 44661  
||| 1797 CTCCCATGAATGATGAATGTTCTTATATGACCTGGAATGGAATCCCTTTCCAAAGGT 1856  
||| 44662 CTCTCATGATCACTAATCTTCTTAATGAGCATGATGAGTGAATCTTTTCAAGAGGT 44721  
||| 1857 TTTCAATTTACTTAAGTCCAGATTCATCCACAGAGATTCACCTTCANAGCAGTTATA 1916  
||| 44722 TTTCAATTTACTTTGCCAGATTCAT-----CAGAGAAATCACCATCTATGACAGGTATA 44776  
||| 1917 GCTTATGGAATGATTTTCTTCAATATAAGGCTTGAAGTGAATTAATCTCTTGATCC 1976  
||| 44777 GCCCTATGAATGATTTTCTTAAATATATAGCTTGAAGTAAATTAATTTCTTGATCC 44835  
||| 1977 ATTTTCTGCAAAATAGATGTTGTG 2000  
||| 44836 ATGGGCTGCAAGATGATGTTGTGAG 44859

RESULT 28  
US-09-949-016-13214  
; Sequence 13214, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: PaSeq for Windows Version 4.0  
; SEQ ID NO 13214  
; LENGTH: 113701  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13214

Query Match 21.7%; Score 433.6; DB 3; Length 113701;  
Best Local Similarity 72.5%; Pired. No. 1.2e-79;  
Matches 760; Conservative 0; Mismatches 229; Indels 59; Gaps 13;

QY 977 TACAGCATACCTTTGAGATGATCTGTGGTTGGTTCATACCCCAATTAATTAACAATA 1036  
||| 6984 TACAGCATACCTTTGAGATGATGTTGGTTCATACCCCA----- 7031

QY 1037 TGCAGAGTGTATATACAAATTAAGTATGACACAGTCTTTTGGCTTCCAGTAT 1096  
||| 7032 ---AAGAAATGAATATTTGAAATACGACGATCACAGGTTTTTGGTTTCTATGTGAT 7088  
||| 1097 ATAAAGTTTGTCTATCTACACTGTAGTCTGTAAAGTATAGTATGCTTAA 1156  
||| 7089 ATAAAGTTTATGTTTATCTATCTATGTTTGTAAATGTCATATAGATCTATCTAA 7148  
||| 1157 AAAAA-CACATACCTTAATTTTAAATGCTTATTAATAAAAAATGCTAACATCATTT 1214  
||| 7149 GAAAAGTATACATCTTATATGTAAGAAATACCTTTTGTCTAAAAAATGCTAACATCATTT 7208  
||| 1215 GAGCATGAGTATGATTTTGTCTGTGTGAGAGTCTTTCTTATATGATACGTA 1274  
||| 7209 GTGCTTACATGAGTATGATGTTTGTCTGTGAGAGTCTTTGTCTGATGATGATGAGC 7268  
||| 1275 T-----CGGGGTCAGTGTCTGAGGCTTGAAGTGTGAGGCTTGTGAGGAGTTCTTAA- 1322  
||| 7269 TGTCTATGATCCGAATGATGTTGTCGCAAGCTGAGGCTGTGTGCAATTTATTAAG 7328  
||| 1323 ---ACAAAGTGAAGATTTGAATATGATGATCTTTCTTCAATGAAGATTTCTCTC 1378  
||| 7329 TAAAGCAAGAGAGATTTGTACATATGATGATGATCTTTCTTCAAG-AAATTTATATA 7387  
||| 1379 TAGGTGTGATGCTTTTGTATGATGATTTTATGACAGTATGAACCTTTGAAATTTG- 1437  
||| 7388 TAGCTGTAAATGCTTTTGTATGATGATTTGTGTGATGAATGAACCTTTTCAATATTTGAA 7447  
||| 1438 TCAATCTCTCAAAACCTGCTGTCTTTAACAACCTAAGTATTAATTAATCTGAATCC 1497  
||| 7448 TCAATCTCTCAAGCTCTGACCTGCTTTATGATGATTAATTTATATA-ATTCTAATTC 7505  
||| 1498 ATTGTGTCAATTTCAACAATTTTCAAGTCTTCAACAGAGATGATTCATCTCATTT 1557  
||| 7506 TTTGTGTCAATTTTACAGAGTGTGTGAACATCTTCACTGAGGATGATTCATCTCAAGA 7565  
||| 1558 CCTGAGATGGAATCTTTGTCTATCCATGAAGAAATTCCTATCTGTCAAGTTTATTC 1617  
||| 7566 -----AACCACTCTTGTCTATCCATTAAGAAACACCTTATTCATTCAGTTTATGC 7619  
||| 1618 ATGAGATGACAGCAATACAGTATGCTTCAAGGCTCACTTCA-CTTTTATTTCAAGTTC 1676  
||| 7620 ATGAGAT-----AAATGTATTAATTTATGATTTTATGAGCTTCACTTGTCTTC 7673  
||| 1677 TCTTGTCTTTTACACATCTGTGATTC-CTTCTCAATTTGAAGCTTGAACCTCTCA 1735  
||| 7674 TTTTGTGTCTTACACATCTGACCTTATCTTCTGCACTGACATCTTGAACCTCA 7733  
||| 1736 AGTATTCATGAGGTTGGAATGACCTTCTCAAAATCTGTATATTTATTTTGTG- 1794  
||| 7734 ACTCATTCAGAGGTTGGAATGACATCTTCAAACTCTGTATATGATATTTTGA 7793  
||| 1795 ---ACCTCCATGAATCAATGAATGTTCTTAATGGAACCTGGAATGGAATCTTTCCAAA 1852  
||| 7794 CCACTCCATGAATTAAGAAATGTTCTTAATGACATCTGAATGGAATCTTTCCAAA 7853  
||| 1853 AGTTTTCATTTACTTATGTCAGATTCATTCATCCAGAGATTCACCTTTCAATGCACT 1912  
||| 7854 AGTTTTCATTTATTTTGTCTGATTTCA-----ACAAAAATACATCTATGAGCAGC 7908  
||| 1913 TATAGCTTATGATGATTTTCTTCAATTAATTAAGGCTTGAAGTTGAATTAATCTCTTG 1972  
||| 7909 TGTTTTCTTATGATGATTTTCTTAATTAATTAATGACATCTGAAGTCAAAATTAATCTCC 7968  
||| 1973 ATCCATTTTCTGCAAAATAGATGTTGTG 2000  
||| 7969 ATCCACAAAGTGAAGTATGTTGTG 7996

RESULT 29  
US-09-949-016-11926/C  
; Sequence 11926, Application US/09949016  
; Patent No. 6812339



```
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 11926
/ LENGTH: 95750
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(95750)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11926
```

Query Match 21.6%; Score 433; DB 3; Length 95750;

Best Local Similarity 72.9%; Pred. No. 1.6e-79;

Matches 721; Conservative 0; Mismatches 225; Indels 43; Gaps 11;

```
QY 1031 CAATATGCAAGATGATATCATCAATAAAGTACACAAAGCTTTGGCTCCCA 1090
DB 92430 CAACATACATAAATGATATCATATAAGTAATCATACAAAGTTT-ATTTCCA 92372
QY 1091 GTGCATATAAAGTTTGCTTACTACTACTAGTCTGTAAAGTGCATATGTAT 1150
DB 92371 GTGCATATAAAGTCTATGTTCCACTATAGTCTAGTCTAGTGTGAGACAGCAT 92312
QY 1151 GTCTAAAAAAC-----ACATACCTTAATTTTAAATGCTTATTACTAAAAATGCTAA 1205
DB 92311 GTTAAACAGAAACAAATGATATCATCTTAATTTTAAATGCTTAAATGCTTAAAAA----- 92258
QY 1206 CAATCATTTGAGCATTCAGTGTATGTAATCTTTTGTGTGGAAGTCTTTTCTTAT 1265
DB 92257 -AAATGCTGTCTTCTTCAATGATCATTAATCTTTTGTGTGATAGAGATCTTGCTCAAT 92199
QY 1266 GATGACTGATCGGGGCTCA-----GGTCTGAAGCTTAAAGGCTGTGTGCACT 1314
DB 92198 GTTGGTGACTGCTGATCTGATCTTAAGTGTAAAGGCTGGGGTGAAGCTGTGACAT 92139
QY 1315 TTCTTAAACAAACAGTGAAGATTTGCAATATCAGTTGACCTTCTTCAATGAAGATTT 1374
DB 92138 TTCTTAAACAAACAGTGAAGATTTGCAATATCAGTTGACCTTCTTCAATGAAGATTT 92079
QY 1375 TCTCTAGTGTGATGCTTTTGAATGATGATTTTATGACAGATGAAGACTCTTTGAAT 1434
DB 92078 TATGTATGATATGAGCTGTGTTGATGATTTTATCCCAATATTAATCTTTTCAAT 92019
QY 1435 GGA-TCAATCTCTCAACCTGCTGTGTTAAACAACCTAATTAATATTTCTGA 1493
DB 92018 GGAGCAATCTTTCAATCTGCTGTGCTTCTCAAGTAAGTTCAATCATATTTGA 91959
QY 1494 ATCCATGTTGTCATTTGCAAAATTTTCAAGTGTTCACAGAGATGATTTCCATCTC 1553
DB 91958 ATCATTTGTTGATTTTAAACATGTTTCAAGGCTTCAACAGATGATCATTTCA 91902
QY 1554 ATTTCCTGAGATGATCTTTGCTCATTCATAGAAGAAATTCCTCATCTGTTCAGTT 1613
DB 91901 -TGATGGGAACACGCTCTTTGCTCATTCATAGAAGAAATTCCTCATCTGTTCAGTT 91843
QY 1614 TATCATGATTTGCAAGCAATACATCATCTTCAAGGCTCACTTCAATTAATTCAG 1673
DB 91842 TATCAGATGATCAGAAATTCAG-CACACCTTCAGG-----CTCCGCTTTCATCTT 91789
```

```
QY 1674 TTCTCTGCTGTTTACACATCTGTGCTCTTCTTCATGATGATCTGACCTCTC 1733
DB 91788 TTCTCTGCTGTTTACACATCTGTGATGATTTCTTCA-AGAAGCTTGAACCTCTC 91730
QY 1734 CAAGTCATCATGAGGTTGGAATGACCTTTCCAAATTCCTGTTAATTAATTTT 1793
DB 91729 AAGTCATCATGAGGATGGAATTAACCTTCCAAATTCCTGTTAATGTTATTTT 91670
QY 1794 GA--CCTCCAGTAATGATGTTCTTAAGGACCTGGAATGATGATCTTCCAA 1851
DB 91669 GACTCTTCCAGTAATGATGTTCTTAAGGATGTTAATGATGATCTTCCAG 91610
QY 1852 AAGTTTCAATTTACTTGTGCAAGTCCATCATCAGAGATCCATTTCAATGCCAG 1911
DB 91609 AAGTTTCAATTTACTTGTGCAAGATCA-----CCAAAGGATCAGATCATGCGAG 91555
QY 1912 TTATAGCTTATGATGATGATTTTCTTAATTAATGAGCTTGAAGTTGAATTTCTCT 1971
DB 91554 CTACAGCTTACCAAAAGATTTCTTAAGTATTAAGACTTGAAGTTAAGATTAATCTCAT 91495
QY 1972 GATCCATTTTCCGCAAAATGATGTTGTG 2000
DB 91494 GATCCATGAGCTGCAAGATGATGCTGTG 91466
```

RESULT 30

US-09-949-016-13144/C

/ Sequence 13144, Application US/09949016

/ Patent No. 6812339

/ GENERAL INFORMATION:

/ APPLICANT: VENTER, J. Craig et al.

/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

/ FILE REFERENCE: CL001307

/ CURRENT APPLICATION NUMBER: US/09/949,016

/ CURRENT FILING DATE: 2000-04-14

/ PRIOR APPLICATION NUMBER: 60/241,755

/ PRIOR FILING DATE: 2000-10-20

/ PRIOR APPLICATION NUMBER: 60/237,768

/ PRIOR FILING DATE: 2000-10-03

/ PRIOR APPLICATION NUMBER: 60/231,498

/ PRIOR FILING DATE: 2000-09-08

/ NUMBER OF SEQ ID NOS: 207012

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO: 13144

/ LENGTH: 112239

/ TYPE: DNA

/ ORGANISM: Human

/ US-09-949-016-13144

Query Match 21.5%; Score 430.8; DB 3; Length 112239;

Best Local Similarity 74.4%; Pred. No. 4.6e-79;

Matches 729; Conservative 0; Mismatches 207; Indels 44; Gaps 13;

```
QY 1047 GGATATCAATAAAGTGAAGTACACAAAGCTTTGGCTTCCAGATGATTAAGTT 1106
DB 81868 GGCCACCAACAAATGAAGTATGAATATTTGTTTCCAGTACATTAAGTTA 81809
QY 1107 TCTTATTAATCAAGTGTGTTAAGTGTGCAATAGTATGATCTTAAAAAAC----- 1162
DB 81808 TCTTATTAATCAAGTGTGTTAAGTGTGCAATAGTATGATCTTAAAAAAC----- 81750
QY 1163 -ACATACCTTAATTTTAAATGCTTTATTAATTAATTAATGCTTAACATCTTGA 1221
DB 81749 TACATACCTTAATTTTAAATGCTTTATTTGTTT-AAAATGCTTAAGATCATCTGAGCTT 81692
QY 1222 CAGTATGTTATCTTTTGTGTTGTTGGAAGTCTTTCTTATTAATGATGAGTGA----- 1274
DB 81691 TGTATGATCATTAATCTTTTGTGTTGTTGGAAGTCTTTCTTATTAATGATGAGTGA----- 81632
QY 1275 ----TCGGGGGTCAAGTGTGTAAGTGTGAGTGTGAGTGTGAGTGTGTAATTA 1328
DB 81631 TGATTAAGGTGTGATTTGCCGAAGCCGAGGTGATGATGAGTGTGAGTGTGTAATTA 81572
```

1329 GTGAAGATTGCAATATCAGTTGACTCTTCTTTCAGTGAAGATTTCTCTAGTGTGA 1388  
1571 ATGAAGTTTGTGCATCATATGACTCTTCTTCCGACAGATTTCTGTGATGATGA 81512  
1389 TGTCTTTGATRGACATTTTATGACAGATGAATCTTCTTGAATTTGA-TCATCTCT 1447  
81511 TGTGTGTGATGACATTTTATGACATGATGAATCTTCTTGAATTTGACATCTCTG 81452  
1448 CAATCCGCTGCTGCTTTAACAACCTAATGATTAATTAATCTGATCATGTTGTGA 1507  
81451 CAATCCGCTGCTGCTTTAACAACCTAATGATTAATTAATCTGATCATGTTGTGA 81333  
1508 TTTCAACATTTTTCACAGTGTCTTCAACAGAGATGATTCATCTCATTTCTGAGATG 1567  
81332 TCTCAACATGTTTGCACATCTTCAATGAGATGATTTCAATCTCAAGAACACTT-- 81335  
1568 AATCTTTGCTCCATGAAGAAATTTCTCATCTGTTCAAGTTTATCATGATGTGC 1627  
81334 --TCTTACTCTTCATAGGAAGCAACTCTCATTCATTAAGTTTATTAAGAGTTGT 81277  
1628 AGCATATGACATGCTTCAAGGCTCATCTCATTTTAATTCAGTTCTGCTGCTT 1687  
81276 AGCATATGACATGCTTCAAGGCTCATCTCATTTTAATTCAGTTCTGCTGCTT 81222  
1688 CTACCAACATCTGCTTCTCTCTCATTTGAAGTCTTGAACCTCTCAAGTATCATGA 1747  
81221 CCACCAACATCTGCTTCTCTCTCATTTGAAGTCTTGAACCTCTCAAGTATCATGA 81162  
1748 GGGTTGAATGACATCTTCTCAAAATCTGTTTAATTTATTTATTTGACCTC--CATG 1804  
81161 GGGCTGATGACATCTTCTCAAACTCCATTAATTTATTTATTTTACTCTTCTCAAG 81102  
1805 AATCATGATGTTCTTAATGGA--CTGGAATGTTGAATCTTCTCAAAAGTTTGA 1861  
81101 AATCATGATGTTCTTAATGGA--CTGGAATGTTGAATCTTCTCAAAAGTTTGA 81042  
1862 ATTTCTTATGTCAGATTCATTCATGAGATTCATCTTCAATCCAGTTATATGCTT 1921  
81041 ATTTCTTATGTCAGATTCATTCATGAGATTCATCTTCAATCCAGTTATATGCTT 80987  
1922 ATGAATGATTTCTTCAATTAATAGGCTTG--AAAGTTGAATTTACTCTTGAATCATTT 1980  
80986 ATGAATGATTTCTTCAATTAATAGGCTTG--AAAGTTGAATTTACTCTTGAATCATTT 80927  
1981 TCTGCAATATGATGTTTG 2000  
80926 GCTACGAAGGATGTTTG 80907

RESULT 31  
US-09-949-016-13498/C  
Sequence 13498, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001037  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: PadsSeq for Windows Version 4.0  
SEQ ID NO 13498  
LENGTH: 343352  
TYPE: DNA  
ORGANISM: Human

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(343352)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13498

Query Match 21.4%; Score 429; DB 3; Length 343352;  
Best Local Similarity 72.7%; Pred. No. 1.4e-78;  
Matches 774; Conservative 0; Mismatches 215; Indels 76; Gaps 14;

977 TACAGGATTCATCTGAGATATCTGAGTTGGTTGGTTCCATACCAACATATATACAAATA 1036  
251976 TACAGGATTCATCTGAGATATCTGAGTTGGTTGGTTCCATACCAACATATATACAAATA 251926  
1037 TCAAGAGTGTATTCACAAATTAAGTACACAGAGTCTTTTGGCTCCAGTGCAT 1096  
251925 -----AGTGAATACCAATTAATTAAGTACACAAATATTTTGGTTCCAGTGCAT 251872  
1097 ATAAAGTTTGGCTTATCTACAGTGTGCTGTTGAAGTGCATATGTTATGTTTAA 1156  
251871 ACAAAAGCTATGTTTACAGTATCTGATGCTATTAAGTGTGCATATACACTAGTTTAA 251812  
1157 AAAA-----CAGTACCTTATTTTAAATGCTTTATTAATAAATGCTAACAT 1210  
251811 ATAAAGCAATGCAATATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 251752  
1211 ATTGAGATTCAGTGAATGTTATCTTTTGGCTGTTGAAGTCTTTTATTTATGATGA 1270  
251751 ATTGAGATTCAGTGAATGTTATCTTTTGGCTGTTGAAGTCTTTTATTTATGATGA 251692  
1271 CTGAT-----CGGGGTCAAGTGTGTAAGTCTTGAAGTGTGCTGCTGCAAGTTTCT 1318  
251691 TGGTGTGACCCGATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251632  
1319 -----TAAACACAGTGAAGATTTGCAATATGATGATGATGATGATGATGATGATGAT 1374  
251631 GAAATTAACACAGTGAAGATTTGCAATATGATGATGATGATGATGATGATGATGATGAT 251572  
1375 TCTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1434  
251571 TCTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251512  
1435 GGA-TCATCTCTGCAACCTGCTGCTTTAA-----CACTAAGTTATTAATA 1487  
251511 AAGTCAATCTCTGCAACCTGCTGCTTTAA-----CACTAAGTTATTAATA 251452  
1488 TTCTGAATCCAT-----TGTTCATTTCAACAAATTTCAACAGTGTCTTCACCA 1536  
251451 TTCTGAATCTCTGCTGCTGCTTTAA-----TGTTCATTTCAACAAATTTCAACAGTGTCTTCACCA 251392  
1537 GGAGTATTCATCTCATTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596  
251391 GGAGTATTCATCTCATTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 251336  
1597 CTCATCTGTTCAAGTTTATTCATGAGATTCAGCAATTAAGTGTCTTCAAGGCTTCA 1656  
251335 CTCATCTGTTCAAGTTTATTCATGAGATTCAGCAATTAAGTGTCTTCAAGGCTTCA 251281  
1657 TTCATCTTAAATTCAGTCTCTGCTGTTTCAACAAATTCAGTGTCTTCTCTCATTT 1716  
251280 TCCAGTCTGATTTCACTTCTCTGCTGTTTCAACAAATTCAGTGTCTTCTCTCATTT 251222  
1717 GAAGTGTGAACCTCTCAAGTATTCATGAGGTTGAATGATGATGATGATGATGATGATGAT 1776  
251221 GAAGGCTTGA--TTCACAAATTCATGAGGTTGAATGATGATGATGATGATGATGATGAT 251164  
1777 GTTAATATTTAATTTTGA-----CTTCCATGATGATGATGATGATGATGATGATGATGAT 1833  
251163 GTTAATATTTGATTTGATCTTCTCCATGATGATGATGATGATGATGATGATGATGATGAT 251104  
1834 ATGCTGAATCTTTCCAAAAGTTTCAATTTACTTATGTCAGATGATGATGATGATGATGAT 1893  
251103 ATAGTGTATTTTCCGGAAGTTTCAATTTACTTATGAT-----CAAGCAGAAC 251049



Best Local Similarity 73.3%; Pred. No. 1.8e-78;  
Matches 762; Conservative 0; Mismatches 211; Indels 67; Gaps 14;

```
OY 978 ACAGGACATCCCTTGGAGATCTGCTGGTTCCTCAACCCGCAATTAATACAAATAT 1037
Db 137722 ACAGGACATCCCTTGGAGATCTGCTGGTTCCTCAACCCGCAATTAATACAAATAT 137781
OY 1038 GCAGAGAGTGGATATCACAATTAAGTGAATCAGACAGTCTTTGGCTTCCAGTGACATA 1097
Db 137782 -----TTGCGATGAATGAGTCAACAAATGTTTTTTGTTCCGGGTCATA 137827
OY 1098 TAAAGTTTGTCTTATCTACTACCTGAGTCTGTTAAGTGTGAATAGCTTATGCTTAA 1157
Db 137828 TGAAGTTATGCTATCTATCTAGTCTGTTAAGTGTCCAAATAGCATATGCTTAA 137887
OY 1158 AAAACACATCCTTATTTTAAATGCTTATTAACAAAAATGCTAACAAATTTGAG 1217
Db 137888 AAATGACATACCTTAAATTAATAAATCTTTATTAAT-AAAAATGCTAATGATCATCTGAG 137946
OY 1218 CATTCAGTGAAGTGAATCTTTTGTGCTGGTGAAGGCTTTTCTTATGATGATGAT- 1275
Db 137947 CTTTACGACAGTGCATATATCTTTTGTGCTGAGGGCTTTGCTTGAATGTTATGGCTGC 138006
OY 1276 -----CGGGGCTCAGGTGCTGAAGCTTAAGGCTGCTGCTGCTGCTGCTTAA 1323
Db 138007 TGACCGATCACAGTATGCTGCTGCCGAAGGCTGGGGTGGCGCTGGCAATTTCTTAA 138066
OY 1334 CAACAGTGAAGATTCGAATATGATGATCTTCTTATGATGAAGATTTCTCTGAGTG 1383
Db 138067 CAA---GGAAAGTTTGGACATATGATATCTTCTTATGATGAAGATTTCTCTGAGTA 138122
OY 1384 TGTGATGCTTTTGAATGATTTTATGACAGTGAAGATCTTCTTGAATTTGA-TCAT 1442
Db 138123 TGGGATGCTGTTTGAACCGCATTTTACCCAGTGAAGATCTTCTCTCAAAATTTGAGTCAAT 138182
OY 1443 CCTCTCAAACTCTGCTCTCTTAAACAACCTTAATTAATTAATTTGATTCATTTGT 1502
Db 138183 CCTCTCAAACTCTGCTCTCTTAAACAACCTTAATTAATTAATTTGATTCATTTGT 138242
OY 1503 TGTGATTTAAACAATTTTACAGTGTCTTCAACAGAGATGAATTCATCTCATTTCTCTA 1562
Db 138243 T-TCAATTTCAACA-----GTTTCAACAGAGATGAATTCATCTCATTTCTCTA 138290
OY 1563 GATGGAATCTTTGCTCATCATTAAGAAGAAATTCCTCATCTGTTCAAGTTTATCATGAG 1622
Db 138291 CTT-----TCTTGTCTATCTCAATGAAGCAACCCCTCATTTATTAAGTTTATCTTGA 138346
OY 1623 ATTGACGAATACAGTATGCTTCAAGGCTCACTTCACTTTTAATTCAGATTCCTCTGC 1682
Db 138347 ATTGACGAATTCAGTATCTTCACTTCAATCTTCACTTC-----TAATATCTTATCTTTC 138400
OY 1683 TGTGTTTACCAATCTGTGCTTCTTCCATTTGAAGTCTTGAACCTTCCCAAGTCAATC 1742
Db 138401 TATTTTCCACATGTCTCAAGTGTCTTCTTCTTCTGAGTCTTGAACCCCTCAAGTCAATC 138460
OY 1743 CATGAGGTTTGAATGAGCTTCTTCAAAATTCCTGTTAATTAATTAATTTGAC---CTC 1799
Db 138461 CATGAGGTTTGAATGAGCTTCTTCAAACTCTGTTAATTAATTTGACCTTTCTT 138520
OY 1800 CCATGATCATGATGTTCTTAAATGAGCACTGGAATGAGTGAATCTTTTCCAAAAGTTT 1859
Db 138521 TATATGATCACAATATGTTCTTAATGATCATCAAGATGAGTGAATTC-CTCCCAAGGTTT 138579
OY 1860 CAATTTACTTATGTCAGATCCATTCACAGAGATCCATTTCAATGCTCAATTAATGCC 1919
Db 138580 CAATTTACTTATGTCAGATCTGT-----AAGAGGAATATGCTATCTAATGAGAGCTGAGAC 138634
OY 1920 TTAATGAATGATTTTCTGCTAATAATAGC--TTGAAGTTGAATTTACTCTTGAATCCA 1977
Db 138635 TTAATGAATGATTTTCTGCTAATAATGAGCTTTTCAAGTCAAAATTAATCTTTGATTTCA 138694
OY 1978 TTTTCTGCAAAATAGATGTT 1997
```

Db 138695 TGGGCTGCAGATGATATTT 138714

```
RESULT 34
US-09-949-016-16419
; Sequence 16419, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 16419
; LENGTH: 61461
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16419
```

Query Match 21.4%; Score 427.8; DB 3; Length 61461;  
Best Local Similarity 73.8%; Pred. No. 1.7e-78;  
Matches 717; Conservative 0; Mismatches 212; Indels 42; Gaps 12;

```
OY 1051 ATCAATATAAGTGAAGTGCACACAGTCTTTGGCTTCCAGTCAATATAAAGTTTGT 1110
Db 39855 ACCACAAATTAAGCAAGTGCACATTAATTTTGTCTTCCAGTCAATATAAAGTTATGT 39914
OY 1111 TATCTACCTGCTAGTCTGTTAAGTGTGCAATAGTCTATGCTTAAATA---ACACT 1166
Db 39915 TTTGCTAATCTGCTAGTCTAAT--TTGTGACACAGCATTAATGCTTAAATAAACAATGTAT 39972
OY 1167 ACCCTAATTTTAAATGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1226
Db 39973 ACCCTAATTTTAAATGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40032
OY 1227 AGTTGAATCTTTTGTCTGTTGAAGGCTTTTC-----TTATGATGACTGATC 1276
Db 40033 AGTTGAATCTTTTGTCTGTTGAAGGCTTTTCCTTAAGGCTGATGATGATGATGATGATG 40092
OY 1277 GGGGGTCAAGT--GCTGAAGCTTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1332
Db 40093 AGGGTATGCTGCTGAAGGCTTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 40152
OY 1333 AGATTGCAATATCATGTTGATCTCTTCTTCAATGAAGATTTCTCTAGTGTGATGCT 1392
Db 40153 AGTTGATGATCATGATTAATCTTTCTTCAAGAAAGATTTCTGTTACATGAAGAGCT 40212
OY 1393 TTTTGTATGATTTTATGACAGTGAAGCTTCTTGAATTAATGAATCAATCTCTCAAC 1452
Db 40213 ATTTGATGATTTTATACCAAGTACATCTTTTCAACATTTGAAGTCAATCTCTCAAA- 40271
OY 1453 CCTGCTCTGCTTAAACAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1512
Db 40272 -CTGCACTGCTTTATCTGTAAGATTTGTGAATATTTCAAAATCTTGTGTATTTTCA 40330
OY 1513 ACAATTTTCAAGTGTCTTCAACAGAGATGAATTCATCTCATTTCTGATGATGATGAT 1572
Db 40331 GCAATGTTCAAGCATCTTCAACAGAGATGATTCATCTCAAGAAACCACTT---TCT 40386
OY 1573 TTGCTCATCATTAAGAAGAAATTCCTCATCTGTTCAAGTTTATCATGATTTGACAGCA 1632
Db 40387 TTGCTCAT-CATTAAGAAGAAATCTTTCATTTCAATTAAGTTTATCATGATTTGACAGCA 40445
OY 1633 TACAGTATGCTTCTTCAAGGCTCACTTCACTTTAATTCAGATTCCTTGTCTTCTACC 1692
```

Db 40446 TTATGACAGGCTTACAGCTCCACATC-----TAATGCGTATCTTACTGTTCTGCC 40499  
Qy 1693 ACATCGTGTGCTTCCCTCCCAATGTAAGTCTGGAACCTCCCAAGCATCCATGAGGTT 1752  
Db 40500 ACAGCAGCAGTGACTTCTCCACCATATGCTTGAACCTCCCAAGCATCCATGAGGTT 40559  
Qy 1753 GGAATCGACTCTTCCCAATTCCTGTTAATATTATTTTGA---CTCCCATGATCA 1809  
Db 40560 GGAATCAACATCTTCCCAATTCCTGTTAATATTATTTGA---CTCCCATGATCA 40619  
Qy 1810 TGAATGTTCTTAATGAGCCTTGAATGTGAATCTTCCCAAGTTTCAATTTACT 1869  
Db 40620 CAAATGTTCTTAATGAGCCTTGAATGTGAATCTTCCCAAGTTTCAATTTACT 40678  
Qy 1870 ACTCCAGATCCATCCATCCAGAGATCCATCTTCAATGCCATTTAGCTTATGGAATG 1929  
Db 40679 TCCTCTTATTCAT-----CAGGAATCATTTAATAGCAGCTATTAATTAATA 40733  
Qy 1930 TATTTCTTCAATTAATGAGCTTGAAGTTGAATTAATCTGATTCATTTCTGCAAA 1989  
Db 40734 TATTTCTTCAATTAATGAGCTTGAAGTTGAATTAATCTGATTCATGAGCTATGAG 40793  
Qy 1990 TAGATGTTGTG 2000  
Db 40794 TGAAGTTGTG 40804

RESULT 35  
US-09-949-016-14546/c  
; Sequence 14546, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FaalSeq for Windows Version 4.0  
; SEQ ID NO 14546  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14546

Query Match 21.3%; Score 426; DB 3; Length 818128;  
Best Local Similarity 74.0%; Pred. No. 6,8e-78;  
Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;  
Qy 1043 AAGTGATATTCATCAATAAGTAGTACACAAAGTCTTTGGCTTCCAGTCATTAATA 1102  
Db 193812 AAGTAAATATTCATCAATAAATAGTACACAAATATTTTGGTCCCGAGTCATTAAT 193753  
Qy 1103 GTTTTGCTATCTACATGTAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGA 1161  
Db 193752 GTTATGTTTACACTCATCATATCTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGA 193693  
Qy 1162 CACATACCTTAATTTTAAATGCTTTTATTAAT-AAAAATCTTAACAATCAATTTGAGCA 1219  
Db 193692 TACATACCTTAATTTTAAATGCTTTTATTAAT-AAAAATCTTAACAATCAATTTGAGCC 193633

Qy 1220 TTCAGTGAAGTTGAATCTTTTGTGCTGAGGAAAGTCTTTTCTTATGATGATGAT----- 1275  
Db 193632 TGCAGGAGTCAATATCTTTTGTGCTGAGGAAAGTCTTTTCTTATGATGATGAT----- 193573  
Qy 1276 -----CGGGGCTAGTGTGCTGAGGAAAGTCTTTTGTGCTGAGGAAAGTCTTTTGTGA 1323  
Db 193572 ATTTGATCAAGGCTGAGGCTGCTGAGGAAAGTCTTTGAGGAAAGTCTTTGAAATTAAGA 193513  
Qy 1324 CAAAGTGAAGTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1382  
Db 193512 CAAAGTGAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 193453  
Qy 1383 GTGTATGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1441  
Db 193452 ATGTATGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 193395  
Qy 1442 TCCTCTCAAAACCCTGCTGCTTTTCAACCACTTAATTAATTAATTAATTAATTAATTAAT 1501  
Db 193394 TCCTCTCAAAACCCTGATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 193335  
Qy 1502 TTGTCAATTTCAACATTTTCAAGTGTCTTCAACAGAGTGAATTCATCTCATTTTCTG 1561  
Db 193334 TTGTCAATTTCAACATTTTCAAGTGTCTTCAACAGAGTGAATTCATCTCATTTTCTG 193275  
Qy 1562 AGATGGAATCTTGTGCTCATCATTAAGAAATTCCTCATCTGTTCAAGTTTATGAGA 1621  
Db 193274 AC-----ATCTTGTCTCATCATTAAGAAATTCCTCATCTGTTCAAGTTTATGAGA 193219  
Qy 1622 GATTCAGCAATATAGTCAATGCTTCAAGGCTTCACTTCACTTCAATTTTCAAGTCTTCTG 1681  
Db 193218 GATTCAGCAATATAGTCAATGCTTCAAGGCTTCACTTCACTTCAATTTTCAAGTCTTCTG 193164  
Qy 1682 CTGTTTCTTCAACATCTGTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1741  
Db 193163 CTGTTTCTTCAACATCTGTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 193105  
Qy 1742 CCATGAGGTTGGAAATGCACTTCTTCAAAATTCCTGTTAATTAATTAATTTGA---CCT 1798  
Db 193104 TGATGAGGTTGGAAATGCACTTCTTCAAAATTCCTGTTAATTAATTAATTTGA---CCT 193045  
Qy 1799 CCCATGAATCATGAATGTTCTTAAATGAGCACTTGAATGAGTGAATCTTCTTCAAAAGTTT 1858  
Db 193044 CCCATGAATCATGAATGTTCTTAAATGAGCACTTGAATGAGTGAATCTTCTTCAAAAGTTT 192990  
Qy 1859 TCAATTTACTTATGTCAGATTCATTCATTCAGAGATTCATCTTCAATGCCATTAATGAC 1918  
Db 192989 TCAGTTTACTTATGTCAGATTCATTCATTCAGAGATTCATCTTCAATGCCATTAATGAC 192934  
Qy 1919 CTTATGGAATGATTTCTTCAATTAATAGGCTTGAAGTTGAATTAATCTCTGATTCAT 1978  
Db 192933 TTTATGGAATGATTTATGATTAATTAATAGGCTTGAAGTTGAATTAATCTCTGATTCAT 192874  
Qy 1979 TTTCTCAAAATATAGTGTGTG 2000  
Db 192873 GGACTGCAAGATGATGTTGTG 192852

RESULT 36  
US-09-949-016-14547/c  
; Sequence 14547, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498

```

:
: PRIOR FILING DATE: 2000-09-08
:
: NUMBER OF SEO ID NOS: 207012
:
: SOFTWARE: FALTSQ for Windows Version 4.0
:
: SEO ID NO 14547
:
: LENGTH: 818128
:
: TYPE: DNA
:
: ORGANISM: Human
:
: FEATURES:
:
: NAME/KEY: misc feature
:
: LOCATION: (1) _.(818128)
:
: OTHER INFORMATION: n = A,T,C or G
:
: US-09-949-016-14547

```

Query Match	21.3%;	Score 426;	DB 3;	Length 818128;
Best Local Similarity	74.0%;	Pred. No. 6.8e-78;		
Matches 727;	Conservative	0;	Mismatches 210;	Indels 45; Gaps 13

OY	1043	AAGTGGATATCAACATTAAGTAGTCAACAAGTCTTTGGCTCCACAGTCATATAAA	1102
Db	193812	AAGTAAATATCCAAATATAATGAGGACACGAAATATTTTGGTCCCGATGATTAAT	193753
OY	1103	GTTTTGTATTACACACTGTAAGTCTGTTAAAGTGTGCAATAGTGTATGCTTAAAAA	-A 1161
Db	193752	GTATAGTTTACACTCTACTATCATCTGTCAAGTGTGCAATAGCATTTGTTAAAAAG	1936933
OY	1162	CACATACCTTATTTTAAAAATGCTTATTACT--AAAAATGTAAACATCATTTAGCA	1219
Db	193682	TACATACCTTAAATTTGAAATACATTACTCTTAAAAAAATGTAAACATCATTTAGCC	193633
OY	1220	TTCACTGAGTTGTAATCTTTTGTGTGGAAAGTCTTTTCTTATTTGATGACTGAT	---- 1275
Db	193632	TGCAGGAGTCAATAATCTTTTGTGTGTGTCAGGCTCTTGCTCAGTAGGATGGCTGTG	1935733
OY	1276	-----CGGGGCTCAGGTGCTGAGACCTTAGAGGTGCTGTGGATTTCTTAA	-----A 1323
Db	193572	ATTGATCAAGGGTGGCGGTGTGTGAAAGTGGGATGGCTGAGCAATTTCTTAAATAG	193513
OY	1324	CAACGTGAAGATTTGCATATATCAGTTGACTCTTCCCTT-TCATGAAATTTCTCTAGT	1382
Db	193512	CAACCAAGAAATTTGCTGCATGATGGCTCTCTCTTATTAGCCAAATGATTTCTTTAGC	1934533
OY	1383	GTGTGATGCTTTTGTGATAGCAATTTTATGCAAGTAGAATCTTTGAAAAATTGGA	-TCAA 1441
Db	193452	ATGTGATGCTGTTTGTATGATCATTTTA--CTCAGTAGAATCTTTCAAAATTTGAGTCA	1933955
OY	1442	TCCTCTCAAAACCTGCTGCTTTTAAACAACTTAAGTTAATATATATTTGCAATTC	1501
Db	193394	TCCTTCCAAACCTGTAGTCAATTTTATCACTTAAGTTATGAGAAATTTCAAAATCTTTG	1933355
OY	1502	TTGTCAATTCACAAATTTTCAACAGTGTCTTCAACAGAGTAGAATTCATCTCATTTCC	1561
Db	193334	TTGTCAATTTCAACAAATTTTCAACAGCATCATACCAAGATAGAGTGCATCTCATGAAC	1932757
OY	1562	AGATGGAATCTTTGCTCATTCATAGAGAAATTTCTCATCTGTTCAGTTTATCATGA	1621
Db	193274	AC-----AATCTTTGCTCATTCATAGAGCAATCTGGAATCCCGTGAAGTTTAACTTGA	1932189
OY	1622	GATTCAGCAATACATCATGTCTTTCAAGGCTCACTTCACCTTTTAAATCCAGTTCTCTG	1681
Db	193218	GATTCAGGTATTTCAAGTCAACCTT-----CTAATTCACCTTCTCATTTCTGGTTCATG	1931646
OY	1682	CTGTTTCTACACATCTGAGGTCTCTTCCATAGTAAGTCTTGAACCTCTCCAAAGCAT	1741
Db	193163	CTAATTTCTACACATCTGCAAGTACCTTTCTTACTGAAGCTTGGACC-CTCAAGTCAT	1931055
OY	1742	CCATGAGGTTGGAAATCGACTTCTCCAAATTCCTGTTAATATTATATTTTGA--	---CCT 1798
Db	193104	TCATGAGTGTGGAAATTAACCTTCACTTCAAACTCCTGATTAATGTAATTTTGA	CTCCT 1930455
OY	1799	CCCATGAATCATGAATGTTCTTAAATGGACCTGGAATGTAATCTTTTCCAAAAGTTT	1858
Db	193044	CCCTTCATCATGAATTTCTTATAGGCATCTGGAATGGAGAGACCTTTCC-----GTTT	1929905

Qy	1859	CCAATTATCTAGTCCAAATCATCATCAGAGATCCATTTCAATCCAGTTATAG	1918
Qy	192989	TCAGTTTACTTATTCAGATCCATAGT---	GGACTACCAATATAGCAGCTAATACC 1929354
Qy	1919	CTTATGGAATGATTTCTTCATAATAAGGCTTGAAGTTGAATTACTCTTGATCCAT	1978
Qy	192933	TTTATGAATGATGATTTAAAGTAAATTAAGACTTGAAGTGAAGTCACTTTGATCCAA	192874
Qy	1979	TTTTCGAAAATAGATGTTGNG	2000
Qy	192873	GGACTGCAGATGATGTTGNG	192852

RESULT 37  
US-09-949-016-14548/c

```

/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO. 14548
/ LENGTH: 818128
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(818128)
/ OTHER INFORMATION: n = A,T,C OR G
US-09-949-016-14548

```

Query Match	21.3%;	Score 426;	DB 3;	Length 818128;
Best Local Similarity	74.0%;	Pred. No. 6.8e-78;		
Matches 727;	Conservative	0;	Mismatches 210;	Indels 45;
				Gaps 13;

QY	1043	AAGTGGATATTCACATTAAGAAGTCACACAAGCTCTTTGGCTCCGAGTGCATTAATAA	1102
Db	193812	AAGTAAATATATCAATTAATAATGAGCACAGAAATATTTTGGTCTCCAGTGCATTAATAAT	193753
QY	1103	GTTTGGCTTAATACACACTGATAGTCTGTTAAGTGCATAAGTGTATGCTCTAAAAAA-A	1161
Db	193752	GTTATGTTTACCTACTACTATCATCTGTCAAGTGTGCATATGACATTGATCTTAAAAAAG	193693
QY	1162	CACATACCTTAATTTTAAATGCTTTATTTACT--AAAAATGCTAACATCAATTGGACA	1219
Db	193692	TACATACCTTAATTTGAAAAATACATTACTCTAAAAAAAATGTAACATCAATCTGAGCC	193633
QY	1220	TTCACTAGAGTTGTATCTTTTTGCGTGGGAAGGCTTTCTTATTTGATGACGTAT----	1275
Db	193632	TGCAGGAGTCAATTAATCTTTTGTCTGGTACAGGCTCTTGCCTCAGTGAAGAAGGCTGCTG	193573
QY	1276	-----CGGGGGTCAGGTCGTGAAGCTTAGGGGTGGCTGTGCAGTTTCTTAAA-----A	1323
Db	193572	ATTGATCAGGGTGGCGGCTTGTCTGAAGGTGGGAATGGCTGTACCAATTTCTTAAAAATAAGA	193513
QY	1324	CAACAGTGAAGATTGCATATATCAGTTGACTCTTCTT--TCATGAAAGATTTCTCTAAT	1382
Db	193512	CAACAAAGAAATTGGCTGCATGCATGATGGCTCTCTCTTATATAGCAATGATTTCTTTGTAC	193453
QY	1383	GGTGTATGCTTTTGTATAGCATTTTATATGACAGTGAACCTCTTGAANAATTGA--TCAA	1441
Db	193452	ATGTATATGCTGTTTATATGACATTTTA--CTCAGTAGAACCTCTTTCAAAATTTGAGTCAA	193395



QY 1442 TCCCTCAACCCCTGCTGCTTTAAACCACTTAAGTTAATATATATTCGATTCATG 1501  
|||  
Db 193394 TCTTCCAAACCTGATGCTAATTTTATCACTAAGTTATGAAATATCTTGG 193335  
|||  
QY 1502 TTGTCAATTCACAAATTTTCAAGTGTCTTACAGAGTAGATTCATCTCATTTCTG 1561  
193334 TTGTCAATTCACAAATTTTCAAGTGTCTTACAGAGTAGATTCATCTCATTTCTG 193275  
|||  
QY 1562 AGATGGAATCTTGTCTCATCCATTAAGAGAAATTCCTCATCTGTCTCAAGTTTATATGA 1621  
193274 AC-----AATCTTGTCTCATCCATTAAGAGAAATTCCTCATCTGTCTCAAGTTTATATGA 193219  
|||  
QY 1622 GATTGAGCAATACAGTCAATGCTTCAAGGCTTCATCTCACTTTTAATTCAGTTCTCTG 1681  
193218 GATTGAGCAATACAGTCAATGCTTCAAGGCTTCATCTCACTTTTAATTCAGTTCTCTG 193164  
|||  
QY 1682 CTGTCTTCAACACATCTGTGCTTCTCTCAATTAAGTGTGAACCTCTCCAGTCAAT 1741  
193163 CTATTCTTCAACACATCTGTGCTTCTCTCAATTAAGTGTGAACCTCTCCAGTCAAT 193105  
|||  
QY 1742 CCATGAGGTTGGAATCGACTCTTCTTCAATTCCTGTAAATTTATATTTGA---CCT 1798  
193104 TGATGAGTGTGGAATTAATCTTCTCAACCTCTGTAAATTTATTTGACCTCTCT 193045  
|||  
QY 1799 CCCATGAATCATGAATGCTTAAATGCACTGGAATGGAATGCTTCCAAAAGTTT 1858  
193044 CCCCATGAATCATGAATGCTTAAATGCACTGGAATGGAATGCTTCCAAAAGTTT 192990  
|||  
QY 1859 TCAATTTACTTATGTCAGATTCATCCATCCAGAGATTCATCTTCAATGCCAGTTATAG 1918  
192989 TCAGTTTACTTATGTCAGATTCATCCATCCAGAGATTCATCTTCAATGCCAGTTATAG 192934  
|||  
QY 1919 CTTATGGAATGATTTCTTCAATATATAGGCTTGAAGTTGAATTAATCTTGTATCAT 1978  
192933 TTTATGGAATGATTTCTTCAATATATAGGCTTGAAGTTGAATTAATCTTGTATCAT 192874  
|||  
QY 1979 TTTCTGCAAAATAGATGTTGTG 2000  
192873 GGACTGCGAATGATGTTGTG 192852  
Db

RESULT 38  
US-09-949-016-14549/c

; Sequence 14549, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIORITY FILING DATE: 2000-04-14  
; PRIORITY APPLICATION NUMBER: 60/241,755  
; PRIORITY FILING DATE: 2000-10-20  
; PRIORITY APPLICATION NUMBER: 60/237,768  
; PRIORITY FILING DATE: 2000-10-03  
; PRIORITY APPLICATION NUMBER: 60/231,498  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14549  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)---(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14549

Query Match 21.3%; Score 426; DB 3; Length 818128;  
Best Local Similarity 74.0%; Pred. No. 6,8e-78;

Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;

QY 1043 AAGTGATATCAATAAAGTAGACACAAGTCTTTGGCTTCCAGTCAATATATAA 1102  
|||  
Db 193812 AAGTAAATATCAATAAATAAGTAGACACAAGATATTTTGGTCCAGTCAATATAT 193753  
|||  
QY 1103 GTTTGGCTTATCTACACTGATGATCTGTAAAGTGTCAATAGTATGTCTTAAATAA-A 1161  
193752 GTTATGTTTACACTGATGATCTGTAAAGTGTCAATAGTATGTATCTTAAATAAAG 193693  
|||  
QY 1162 CACATACCTTAAATTTTAAATAGCTTTATTAAT-AAAAATGTCAATCATTTAGACA 1219  
193692 TACATACCTTAAATTTTAAATAGCTTTATTAAT-AAAAATGTCAATCATTTAGACC 193633  
|||  
QY 1220 TTCAAGTATGTAATCTTTTGTGTGTGTGAAGTCTTTTCTTATGATGATGAT- 1275  
193632 TGCAGGAGTCAATATCTTTTGTGTGTGTGAAGTCTTTTCTTATGATGATGAT- 193573  
|||  
QY 1276 -----CGGGGATCGGTGCTGAAGCTTAAAGGCTGAGGAGTTCTTAA-----A 1323  
193572 ATTTGATCAGGATGAGGCTGAGGCTGAGGAGTGGATGAGTCTTAAATTTCTTAAATAAG 193513  
|||  
QY 1324 CACAGTAAAGATTCATATATCATGTTGACTCTTCTT-TCATGAAAGATTTCTCTAGT 1382  
193512 CACAGTAAAGATTTGCTGATGATGATGCTCTCTTATTAAGCAATGATTTCTTGTAGC 193453  
|||  
QY 1383 GTGTGATGCTTTTGTATGATGATTTATGACAGTGAATCTTTGAAATTTGA-TCGA 1441  
193452 ATGTGATGCTTTTGTATGATGATTTGA-CTCAGTGAATCTTTTCAAAATTTGAGTCA 193395  
|||  
QY 1442 TCCCTCAAAAGCTGCTGCTTTTAAACCAACCTTAAGTTATATATTCGATTCATG 1501  
193394 TCTTCTCAAAAGCTGCTGCTTTTAAATTAATCACTTAAGTTATATTAATTCCTTGTG 193335  
|||  
QY 1502 TTGTCAATTCACAAATTTTCAAGTGTCTTCAACAGAGTAGATTCATCTCATTTCTG 1561  
193334 TTGTCAATTCACAAATTTTCAAGTGTCTTCAACAGAGTAGATTCATCTCATTTCTG 193275  
|||  
QY 1562 AGATGGAATCTTGTCTCATCCATTAAGAGAAATTCCTCATCTGTCTCAAGTTTATATGA 1621  
193274 AC-----AATCTTGTCTCATCCATTAAGAGAAATTCCTCATCTGTCTCAAGTTTATATGA 193219  
|||  
QY 1622 GATTGAGCAATACAGTCAATGCTTCAAGGCTTCATCTCACTTTTAATTCAGTTCTCTG 1681  
193218 GATTGAGCAATACAGTCAATGCTTCAAGGCTTCATCTCACTTTTAATTCAGTTCTCTG 193164  
|||  
QY 1682 CTGTCTTCAACACATCTGTGCTTCTCTCAATTAAGTGTGAACCTCTCCAGTCAAT 1741  
193163 CTATTCTTCAACACATCTGTGCTTCTCTCAATTAAGTGTGAACCTCTCCAGTCAAT 193105  
|||  
QY 1742 CCATGAGGTTGGAATCGACTCTTCTTCAATTCCTGTAAATTTATATTTGA---CCT 1798  
193104 TGATGAGTGTGGAATTAATCTTCTCAACCTCTGTAAATTTATTTGACCTCTCT 193045  
|||  
QY 1799 CCCATGAATCATGAATGCTTAAATGCACTGGAATGGAATGCTTCCAAAAGTTT 1858  
193044 CCCCATGAATCATGAATGCTTAAATGCACTGGAATGGAATGCTTCCAAAAGTTT 192990  
|||  
QY 1859 TCAATTTACTTATGTCAGATTCATCCATCCAGAGATTCATCTTCAATGCCAGTTATAG 1918  
192989 TCAGTTTACTTATGTCAGATTCATCCATCCAGAGATTCATCTTCAATGCCAGTTATAG 192934  
|||  
QY 1919 CTTATGGAATGATTTCTTCAATATATAGGCTTGAAGTTGAATTAATCTTGTATCAT 1978  
192933 TTTATGGAATGATTTCTTCAATATATAGGCTTGAAGTTGAATTAATCTTGTATCAT 192874  
|||  
QY 1979 TTTCTGCAAAATAGATGTTGTG 2000  
192873 GGACTGCGAATGATGTTGTG 192852  
Db

RESULT 39  
US-09-949-016-14550/c

Query Match 21.3%; Score 426; DB 3; Length 818128;  
Best Local Similarity 74.0%; Pred. No. 6,8e-78;



Sequence 14550, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14550  
LENGTH: 818128  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(818128)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14550

Query Match 21.3%; Score 426; DB 3; Length 818128;  
Best Local Similarity 74.0%; Pred. No. 6,8e-78;  
Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;

1043 AAGTGAATATACATTAAGAGTACACACAAAGTCTTTGGCTCCAGTCATATATAA 1102  
193812 AAGTGAATATACATTAAGAGTACACACAAAGTCTTTGGCTCCAGTCATATAT 193753  
1103 GTTTGCTTATACACTGAGTCTGTTAAGTGTCAATAGTGTATGCTTAAAAA-A 1161  
193752 GTTATGTTTACACTGATATATCTGTCAAGTGTCAATAGCATTTGTTAAAAAAG 193693  
1162 CACATACCTTAATTTAAAAAGTCTTATTAAT-AAAAAAGTCAATCAATTTGAGA 1219  
193692 TACATACCTTAATTTAAAAAGTCTTATTAAT-AAAAAAGTCAATCAATTTGAGC 193633  
1220 TTCAGGAGTGTGAATCTTTTGTCTGTGTGAAGGCTTTTCTTATGATGACTGAT- 1275  
193632 TCGAGGAGTGTGAATCTTTTGTCTGTGTGAAGGCTTTTCTTATGATGACTGAT 193573  
1276 -----CGGGGGTCAGGTCCTGAAGCTTAAGGTCGTGTGAGTTCTTAAA-----A 1323  
193572 ATTGATCAGGGTGGCGGTTGCTGAAGTGGGATGGCTGAGCAATTTCTTAAAAATAGA 193513  
1324 CAAAGTGAAGATTCGATATATCAAGTGTCTTCTT-TCATGAAGAATTTCTCTTAAT 1382  
193512 CAAAGTGAAGATTCGATATATCAAGTGTCTTCTT-TCATGAAGAATTTCTCTTAAT 193453  
1383 GTGTGATGCTTTTGAATAGATTTTGAAGCAGTGAAGCTTTCTTGAAGAATTTGA-TCGA 1441  
193452 ATGTGATGCTTTTGAATAGATTTTGAAGCAGTGAAGCTTTCTTGAAGAATTTGA 193395  
1442 TCTCTCAAAACCTGCTGCTTTTAAACAACCTAAGTTAATATATATTTCTGAATTCATG 1501  
193394 TCTCTCAAAACCTGATGCTAATTTTAACTAAGTTAATATATATTTCTGAATTCATG 193335  
1502 TTGTCTTTCACAAATTTTCAAGTGTCTTCAACAGAGTGAATTCATCTCATTTCTG 1561  
193334 TTGTCTTTCACAAATTTTCAAGTGTCTTCAACAGAGTGAATTCATCTCATTTCTG 193275  
1562 AAGTGAATATCTTGTCAATCAGTGTCTTCAACAGAGTGAATTCATCTCATTTCTG 1621  
193274 AC-----AATCTTGTCTATCATAAGAGCAACTGTGATCGCTGAAGTTTAACTTGA 193219  
1622 GATTGAGCAATACAGTATGCTTTCAGGCTCCTCACTTCACTTTTAAATTCAGTTCTCTTG 1681

Db 193218 GATTACAGCTATTCAGTACACCTT-----CTGACTCAGCTTCATCTGTTCTCATG 193164  
Qy 1682 CTGTTTACACACATCTGTGCTCTTCTTCATTAAGTGTGAACCTTCCAGTAT 1741  
Db 193163 CTATTTTACACACATCTGTGCTCTTCTTCATTAAGTGTGAACCTTCCAGTAT 193105  
Qy 1742 CCAATGAGGTTGAATGACCTTCTTCCAAATTTCTGTTAATTTATTTTGA---CCT 1798  
Db 193104 TGATGAGTGTGGAATTAACACTTCCAAACCTCTGATTAAGTTGAATTTTGACTTCT 193045  
Qy 1799 CCCATGAATGATGATGTTCTTAAAGCAGCTGGAATGTGAATCTTCCAAAAGTTT 1858  
Db 193044 CCCCAATGATGATGATGTTCTTAAAGCAGCTGGAATGTGAATCTTCCAAAAGTTT 192990  
Qy 1859 TCAATTTACTTATGTCAGATCCATCCATCCAGAGATCCACTTTCATGCAATTAAGC 1918  
Db 192989 TCAATTTACTTATGTCAGATCCATCCATCCAGAGATCCACTTTCATGCAATTAAGC 192934  
Qy 1919 CTTATGAGATGATTTCTTCAATTAATTAAGCTTGAAGTGAATTTACTCTTGAATTCAT 1978  
Db 192933 TTTATGAATGATGATTTAATTAAGCTTGAAGTGAATTTACTCTTGAATTCAT 192874  
Qy 1979 TTTCTGCAAAATGATGTTGTG 2000  
Db 192873 GGACTGAGAAATGATGTTGTG 192852

RESULT 40  
US-09-949-016-14551/C  
Sequence 14551, Application US/09949016  
Patent No. 6812339

GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14551  
LENGTH: 818128  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(818128)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14551

Query Match 21.3%; Score 426; DB 3; Length 818128;  
Best Local Similarity 74.0%; Pred. No. 6,8e-78;  
Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;

1043 AAGTGAATATACATTAAGAGTACACACAAAGTCTTTGGCTCCAGTCATATATAA 1102  
193812 AAGTGAATATACATTAAGAGTACACACAAAGTCTTTGGCTCCAGTCATATAT 193753  
1103 GTTTGCTTATACACTGAGTCTGTTAAGTGTCAATAGTGTATGCTTAAAAA-A 1161  
193752 GTTATGTTTACACTGATATATCTGTCAAGTGTCAATAGCATTTGTTAAAAAAG 193693  
1162 CACATACCTTAATTTAAAAAGTCTTATTAAT-AAAAAAGTCAATCAATTTGAGA 1219  
193692 TACATACCTTAATTTAAAAAGTCTTATTAAT-AAAAAAGTCAATCAATTTGAGC 193633  
1220 TTCAGGAGTGTGAATCTTTTGTCTGTGTGAAGGCTTTTCTTATGATGACTGAT- 1275

```

Db      193632 TCAGAGGATCATTAATCTTTTGGTGTGACAGGCTCTTGCTCAGAGAGATGGCTGCTG 193573
QY      1276 -----CGGGGGTCAGTGTGCTGAGAGCTTAGGGTGGCTGTGGAGTTTCTTAA-----A 1323
Db      193572 ATTGATCAGGGGTGGCGGGTGTGAGAGGTGGAGATGGCTGTAGCAATTTCTTAAATPAGA 193513
QY      1324 CAACAGTGAAGATTGCAATATCAGTGTGACTTCTCTT-TCATGAAGATTCTCTCTAGT 1382
Db      193512 CAACACGAGAGTTTGTGTGATGATGGCTCTCTCTTTATAGCCAAATGATTTCTTTGTAC 193453
QY      1383 GTGTGATGCTTTTGTATGATGATTTATGACAGATGAGTCTTTGAAATTTGA-TCAA 1441
Db      193452 AGTGTAGTGTGTTGATGATGATTTTA--CTCAGTGAACCTTTCTTCAAAATTTGAGTCA 193395
QY      1442 TCCCTTCACAACTCTGTCTGTCTTAAACAACCTTAAGTATATATATTTGTGATCCATTG 1501
Db      193394 TCTTCTCAAACTCTGTATGCTTAATTTATCAACTAAGTTATGAGATTTCTTAAATCTTTG 193335
QY      1502 TTGTCAATTCACAAATTTTACAGATGCTTACACAGAGTGAATTCATCTCATTTCTG 1561
Db      193334 TTGTCAATTCACAAATTTTACAGATGCTTACACAGAGTGAATTCATCTCATTTCTG 193275
QY      1562 AGATGGAATCTTGTCTCATCATGAGAGAAATTCCTCATCTGTTCAGTTTATCATGA 1621
Db      193274 AC-----AATCTTGTGCTCATCATGAGAGAACTGTGATCGGTGAATTTTACCTTGA 193219
QY      1622 GATTGAGCAATPACAGTCAATGCTTTCAGGCTTCACTTCACTTTAATTCAGTTCTCTG 1681
Db      193218 GATTACAGCTATTCAGTCAACCTT-----CTGACTCCACTTCTCATTCGTGTTCTGAG 193164
QY      1682 CTTTCTTACCAATCTGTGAGTTCCTCTCATTTGAAGTCTTGAACCTCTCCAGTCAAT 1741
Db      193163 CTATTTCTTACCAATCTGTGAGTTCCTCTCATTTGAAGTCTTGAACCTCTCCAGTCAAT 193105
QY      1742 CCATGAGGTTGGAATTCGACTTCTTCCAAATTCCTGTATATTTATTTTGA---CCT 1798
Db      193104 TGATGAGTGTGGAATTAATCACTTCCAAATCTCTGATATATGTTGATTTGACCTCCT 193045
QY      1799 CCCATGATCATGAATGTTCTTAATGCACTGTGAATGTGAAATCTTTCCAAAAGTTT 1858
Db      193044 CCCCTCAATCATGAATTTTCTTATGCGCATGTGAATGTGAGACTTTCC-----GTTT 192990
QY      1859 TCAATTTACTAGTCCAGATTCATGATCCAGAGATTCATCTTCAATGCCAGTTATAC 1918
Db      192989 TCAGTTACTTATCCAGATTCATGAT---GAGCTCACCAATATATGAGGCTATATAC 192934
QY      1919 CTTATGATGATGATTTCTTCAATTAATAGGCTTGAAGTTGAATTTACTCTTGATCCAT 1978
Db      192933 TTTATGAATGATGATTTAAGTATATAGACTTGAAGTTGAAGTCACTTTGATCCAA 192874
QY      1979 TTTCTGCAAAATGATGTTGTG 2000
Db      192873 GGACTGCAGAATGATGTTGTG 192852

```

```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14552
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) --(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552

Query Match      21.3%; Score 426; DB 3; Length 818128;
Best Local Similarity 74.0%; Pred. No. 6,8e-78;
Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;

QY      1043 AAGTGATATCAATTAAGTGAAGTGCACAAAGTCTTTTGGCTTCCAGTGATATAA 1102
Db      193812 AAGTAAATATCAATTAAGTGAAGTGCACAAAGTATTTTGGTCCCGAGTGATTAAT 193753
QY      1103 GTTTGCTTATATCTACCTGTATGCTGTTAAGTGTGCAATATGTTATGCTTAAATAA-A 1161
Db      193752 GTTATGTTTACCTCTATCATCTGTCAAGTGTGCAATATGATGATGATTTAAATAAG 193693
QY      1162 CACATACCTTAATTTTAAATGCTTATATCT-AAAAAATGCTAACATCATTTGAGCA 1219
Db      193692 TACATACCTTAATTTTAAATGCTTATATCT-AAAAAATGCTAACATCATTTGAGCC 193633
QY      1220 TTCAGTGAATGTAATCTTTTGTGCTGTGAGAGGCTTTTCTTATGATGATGAT- 1275
Db      193632 TCAGAGGATCATTAATCTTTTGTGCTGTGAGAGGCTTTTCTTATGATGATGAT- 193573
QY      1276 -----CGGGGGTCAGTGTGCTGAGAGCTTAGGGTGGCTGTGGAGTTTCTTAA-----A 1323
Db      193572 ATTGATCAGGGGTGGCGGGTGTGAGAGGTGGAGATGGCTGTAGCAATTTCTTAAATPAGA 193513
QY      1324 CAACAGTGAAGATTGCAATPACAGTGTGACTTCTCT-TCATGAAGATTCTCTCTAGT 1382
Db      193512 CAACACGAGAGTTTGTGTGATGATGGCTCTCTCTTTATAGCCAAATGATTTCTTTGTAC 193453
QY      1383 GTGTGATGCTTTTGTATGATGATTTATGACAGATGAGTCTTTGAAATTTGA-TCAA 1441
Db      193452 AGTGTAGTGTGTTGATGATGATTTTA--CTCAGTGAACCTTTCTTCAAAATTTGAGTCA 193395
QY      1442 TCCCTTCACAACTCTGTCTGTCTTAAACAACCTTAAGTATATATATTTGTGATCCATTG 1501
Db      193394 TCTTCTCAAACTCTGTATGCTTAATTTATCAACTAAGTTATGAGATTTCTTAAATCTTTG 193335
QY      1502 TTGTCAATTCACAAATTTTACAGATGCTTACACAGAGTGAATTCATCTCATTTCTG 1561
Db      193334 TTGTCAATTCACAAATTTTACAGATGCTTACACAGAGTGAATTCATCTCATTTCTG 193275
QY      1562 AGATGGAATCTTGTCTCATCATGAGAGAAATTCCTCATCTGTTCAGTTTATCATGA 1621
Db      193274 AC-----AATCTTGTGCTCATCATGAGAGAACTGTGATCGGTGAATTTTACCTTGA 193219
QY      1622 GATTGAGCAATPACAGTCAATGCTTTCAGGCTTCACTTCACTTTAATTCAGTTCTCTG 1681
Db      193218 GATTACAGCTATTCAGTCAACCTT-----CTGACTCCACTTCTCATTCGTGTTCTGAG 193164
QY      1682 CTTTCTTACCAATCTGTGAGTTCCTCTCATTTGAAGTCTTGAACCTCTCCAGTCAAT 1741
Db      193163 CTATTTCTTACCAATCTGTGAGTTCCTCTCATTTGAAGTCTTGAACCTCTCCAGTCAAT 193105
QY      1742 CCATGAGGTTGGAATTCGACTTCTTCCAAATTCCTGTATATTTATTTTGA---CCT 1798
Db      193104 TGATGAGTGTGGAATTAATCACTTCCAAATCTCTGATATATGTTGATTTGACCTCCT 193045
QY      1799 CCCATGATCATGAATGTTCTTAAATGCACTGTGAATGTGAAATCTTTCCAAAAGTTT 1858
Db      193044 CCCCTCAATCATGAATTTTCTTATGCGCATGTGAATGTGAGACTTTCC-----GTTT 192990
QY      1859 TCAATTTACTAGTCCAGATTCATGATCCAGAGATTCACATTTCAATGCCAGTTATAGC 1918

```

Db 192989 TCAGTTACTTTATTCACATCCATCAGT-----GGACTCACCAATATATGACGCTATAC 192934  
Qy 1919 CTTATGAATGATTTCTTTCACATATATAGGCTTGAAGTGAATATCTCCGTGATCAT 1978  
Db 192933 TTTATGAATGATGATTTAATTAAGTAATGAAGTGAAGTGAAGTCTTGTATCCA 192874  
Qy 1979 TTTCTGCAAAATAGATGTTGTG 2000  
Db 192873 GGACTGCAAGATGATGTTGTG 192852

RESULT 42  
US-09-949-016-14553/c  
; Sequence 14553, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 14553  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc. feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14553

Query Match 21.3%; Score 426; DB 3; Length 818128;  
Best Local Similarity 74.0%; Pred. No. 6,8e-78;  
Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;

Qy 1043 AAGTGATATCAATTAAGTGAAGTATACCAAGTCTTTGGCTTCCAGTGCATATAAA 1102  
Db 193812 AAGTAAATATCACAATTAAGTGAAGTATATTTTGGCTTCCAGTGCATATAAT 193753  
Qy 1103 GTTTTGCTATATCTACAGTGTGCTGTTAAGTGCATATGATGCTTAATAAAA-A 1161  
Db 193752 GTTAAGTTTACACTCTACTATATCTGTCAAGTGTCAATGATGATGCTTAATAAAAG 193653  
Qy 1162 CACATACCTTAATTTTAAATGCTTTATTAAT-AAAAAATGCTAACATCATTTGAGCA 1219  
Db 193692 TACATACCTTAATTTTAAATGCTTTATTAAT-AAAAAATGCTAACATCATTTGAGCC 193633  
Qy 1220 TTCAAGTGAATGATCTTTTGTGCTGTGGAAGTCTTTTCTTATGATGATGAT- 1275  
Db 193633 TCCAGGAGATCAATCTTTTGTGCTGTGACAGGCTTCTGCTCAAGAGATGCTGCTG 193573  
Qy 1276 -----CGGGGGTCAAGTGCAGTGAAGCTTAAGGCTGCTGAGCTTTCTTAA-----A 1323  
Db 193572 ATTGATCAGGGTGGCGGTGCTGAGAGTGGGATGAGCTGATAGCAATTTCTTAAATAAAG 193513  
Qy 1324 CAACAGTGAAGATTCATATCATGTTGACTCTTCTT-TCATGAAGATTTCTCTAGT 1382  
Db 193512 CAACAGTGAAGATTTGCTGATGATGATGCTCTCTTATAGCAATGATTTCTTTGAGG 193453  
Qy 1383 GTGTGATGCTTTTGTATGATGATTTATGCAAGTGAAGTCTTTGAAAATTGGA-TCGA 1441  
Db 193452 AATGATGCTGTTGATGATGATTTTA--CTCAGTGAAGATCTTTGAAAATTGAGTCAA 193395

Qy 1442 TCTCTCAAACTGCTCTGCTTTTAAACATTAATTAATTAATTTGATTCATATG 1501  
Db 193394 TCTTCTCAAACTGCTCTGCTTTTAAACATTAATTAATTAATTTGATTCATATTC 193335  
Qy 1502 TTGTCATTTCAAAATTTTTCACAGTGTCTTACACAGAGTGAATTCATCTATTTCTG 1561  
Db 193334 TTGTCATTTCAAAATTTTTCACAGTGTCTTACACAGAGTGAATTCATCTATTTCTG 193275  
Qy 1562 AGATGGAATCTTTGCTCATTCATTAAGAAATTTCTTCATCTGTTCAAGTTTATCATGA 1621  
Db 193274 AC-----AATCTTGTCTATTCATTAAGAGCAACTCTGATCGCGAAGTTTACCTGA 193219  
Qy 1622 GATTCAGCAATTAACATCATCTTTCAGGCTCTCATCTTATTCAGTTTCTCTG 1681  
Db 193218 GATTCAGCAATTAACATCATCTTTCAGGCTCTCATCTTATTCAGTTTCTCTG 193164  
Qy 1682 CTGTTTTCACACATCTGCTGCTTCTCATTTGAAGTCTTGAACCTCCAGATCAT 1741  
Db 193163 CTATTTTCACACATCTGCTGCTTCTCATTTGAAGTCTTGAACCTCCAGATCAT 193105  
Qy 1742 CCATGAGGTTGGAATCGACTTCTTCCAAATTTCTGTTAATTAATTAATTTTGA---CCT 1798  
Db 193104 TGATGAGTGTGGAATTTACTACTTCCAACTCCGATTAATGATTTTGAACCTCCT 193045  
Qy 1798 CCCATGAATCATGAATGTTCTTAATGCACTGCAATGATGATCTTTCCAAAGTTT 1858  
Db 193044 CCCCATGAATCATGAATGTTCTTAATGCACTGCAATGATGATCTTTCCAAAGTTT 192990  
Qy 1859 TCAATTTTCTTAATGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1918  
Db 192989 TCAGTTACTTTATTCACATCCATCAGT-----GGACTCACCAATATATGACGCTATAC 192934  
Qy 1919 CTTATGAATGATTTCTTCAATTAATTAAGGCTTGAAGTGAATTAATCTCTGATCAT 1978  
Db 192933 TTTATGAATGATGATTTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 192874  
Qy 1979 TTTCTGCAAAATAGATGTTGTG 2000  
Db 192873 GGACTGCAAGATGATGTTGTG 192852

RESULT 43  
US-09-949-016-14554/c  
; Sequence 14554, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 14554  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc. feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14554

Query Match 21.3%; Score 426; DB 3; Length 818128;  
Best Local Similarity 74.0%; Pred. No. 6,8e-78;  
Matches 727; Conservative 0; Mismatches 210; Indels 45; Gaps 13;

QY	1043	AAGTGGATATATCAACAATAAAGTGATGACACAAAGCTCTTTGGCTTCCCAAGCATATPAAA	1102
Db	193812	AGGTAAATATATCAAAATAAAATGAGGACACAGAAATATTTTGGTTCCCAAGCATPAAAT	1937533
QY	1103	GTTTTGCTTATATCACTAGTACTCTGTAAAGTGCCAAATAGTATTATGTCTAAAAA-A	1161
Db	193752	GTTATGTATACACTTACTATCATCTGTCAAGTGTCAAATAGCATTTGTATCTAAAAAATG	193693
QY	1162	CACATTAACCTTAATTTTAAAAATGTTATTAATCT--AAAAATGCTAACATCATTTGAGCA	1219
Db	193692	TACATAACCTTAATTTGAAAAATATATTATCTTAAAAAAAATGTGTAAACAATCATCTGAGCC	1936333
QY	1220	TTCAAGAGATGTGAATCTTTTGGCTGAGAGAGTCTTTTCTTATTTGATGACTGAT----	1275
Db	193632	TCCAGGAGATCATTAATCTTTTGGCTGTACAGGCTCTTGCTCTAGAGAGATGGCTGCTG	193573
QY	1276	-----CGGGGGATCAGGTGCTGAAAGCTTAAGGTGGCTGTGGCAGTTTCTTAAA-----A	1323
Db	193572	ATTGATCAAGGGTGGCGGTTTGTCTGAAGTGGGATGGCTGTAGCAATTTCTTAAATPAAAG	1935133
QY	1324	CACACTGAAGATGTGCATATATCAGTTGACTCTTCTCTT--TCATGAAGAATTTCTCTTAAT	1382
Db	193512	CACACACGAAGTTTGTGCATGCATGGATGGCTCTCTCTTAAAGCAAAAGATTTCTTGTAGC	1934533
QY	1383	GTGTGATGCTTTTGTATAGCATTTTATGCAAGTAGAAGTCTCTTGTAAAAATGGA-TCGA	1441
Db	193452	ATGTGATGCTGTTGTATAGCATTTTAT--CTCACTAAGACTCTTCTTCAAAAATTTGAGTCA	193395
QY	1442	TCCTCTCAAAACCCTGCTCTGCTTTTAAACAACCTAAGTTAATATATATCTGATTCGATTTG	1501
Db	193394	TCTTCTCAAAACCCTGATGCTAATTTATTCAACTAAGTTTATGGAATATCTTAAATCCTTTG	193335
QY	1502	TTGTCAATTTCAACAATTTTCAAGATCTTTCACAGAGTAGATTCATCTCAATTCCTG	1561
Db	193334	TTGTCAATTTCAACAATTTTCAAGATCTTTCACAGATCTTTCACAGAGTAGATTCATCTCA	193275
QY	1562	AGATGGAATCTTGTGCATCCATGAAGAAAGAAATTCCTCATCTGTCAAGTTTATCATGA	1621
Db	193274	AC-----AATCTTGTCTCATCCATACAGACAGAACTGTGAATCCGCTAAGTTTAACTTGA	193219
QY	1622	GATTCAGCAATACAGATCATGTCTTTCAGGCTTCACCTCACTTCACTTTTAAATTCGAGTTCTT	1681
Db	193218	GATTCAGCTATTCAGTGCACACCTT-----CTGACTCCACTTCTCATTTGGTTCTCAG	193164
QY	1682	CTGTTTCTTACCAATCTGTGTGTTCTTCTCTCCATTTGAAGTCTTGAACCTCTCCAAATCT	1741
Db	193163	CTATTTCTTACCAATCTGTGTGTTCTTCTCTCTTCACTGAAGTCTTGGACC--CTCAAAATCAT	193105
QY	1742	CCATGAGGGTGTGAATCGACTCTTCCCAAAATCCGTGTAAATTTTATTTTGA-----CGT	1798
Db	193104	TGATGATGTGTGAATTAATCTACTTCCAAATCTGTATATTTGTGATTTTGAACCTCTCT	193045
QY	1799	CCCATGAATCATGAATGTTCTTAAATGGCACTGGAATGTGTGAATCTTTCCAAAGGTTT	1858
Db	193044	CCCTTCATCATGAATTTTCTTCAATGGCAGTGTGAATGTGTGAAGGACTTTTCC-----GTTT	192990
QY	1859	TCAATTTTACTTATGTCCAAATCCATCCATCCAGAGATCCACTTTCAATGCCAGTTTATGC	1918
Db	192989	TCAGTTTACTTATTCAGATTCATCTAGT-----GAGCTCACCAATATATGACAGCTTATAC	192934
QY	1919	CTTATGGAATGTATTTCTTCAATTAATPAAAGCTTGAAGTTGAATTAATCTGTGATCCAT	1978
Db	192933	TTTATGAATGTATGTATTTAAGTAAATPAAACCTTGAAGTTGAAGTACTTTGTATCCAA	192874
QY	1979	TTTCTGCAAAATPAGATGTTGTG	2000
Db	192873	GGACTGCAAAATGATGTTGTG	192852

Db	193218	GATTACAGCTATTCAGTACACCTT-----CTGACTCCAACTTCATTCCTGGATTCCTCAGT	193164
Qy	1622	GATTGACGACATTAACGATGATGCTTCCAGGCTCACTTACTTATTTAATTCAGATTCTCTTG	1681
Db	193274	AC-----AATCTTTGGCTATCATCAATAGCAGCAACTGTGATCCGGCTGAAGTTTACCTTGA	193219
Qy	1562	AGATGGAATCTTTGGCTCATTCATAGAAGAAATTCCTCATCTGTTCAAGTTTATCATGA	1621
Db	193334	TTTGTCATTTCAACAATTTTCAACAGTGTCTTCAACCAGAGATGATTCATCTCATTTCCGTG	1561
Qy	1442	TCCTCTCAACCCCTGCTCTGCTTTTAAACAACCTAAGTTAATATATATTTCTGAATCATTTG	1501
Db	193394	TCTTCTCAACCCCTGATGCTAATTTATCAACATTAAGTTATGAATAATTTCTAATATCTTTTG	193345
Qy	1383	GTTGTATGCTTTTGTATGACATTTTATGACACAGTAGAACTTCTTGAATAATTGA-TCNA	1441
Db	193452	ATGTGATGCTGTGTTGATGACATTTTA--CTCAGTAGAACTTCTTCAAAATTTGATGTCNA	193395
Qy	1324	CAACGATGAATGTCATATTCAGTTGACTCTTCTCTT-TCATGAAGAATTTCTCTCTACT	1382
Db	193512	CAACACGAAGTTTCTGTCGATCGATTTGGCTCTCTTTATAGCAATGATTTCTTTGTAGC	193453
Qy	1220	TTCAATGATGTTGTAATCTTTTGTCTGTGTGAAGGCTTTTCTTATGATGACTGAT----	1275
Db	193632	TGCAAGGAGTCATATCTTTTGTCTGTGTACAGGCTCTTGCCCTCACTGAGATGGCTGCTG	193573
Qy	1216	-----CGGGGCTCAGGTCGCTGAAGCTTAAAGGTGGCTGTGCGACTTTCTTAA-----A	1323
Db	193572	ATTGATCAGGGGTGGGTTGTCTGAAGGTCTGGGATGCTGTACCAATTTCTTAAATATAGA	193513
Qy	1162	CACATACCTTATTTTAAATAGCTTTATTTACT--AAAAATGCTAACAATCATTTTGAGCA	1219
Db	193692	TACATACCTTATTTGAAAAATACATATCTCCCTAAAAAATGTAACATATCTGAGGC	193633
Qy	1103	GTTTTGCTTATACACACTGAGTGTGTAAAGTGTGCAATAGTATGTATGCTAAAAAA-A	1161
Db	193812	AAGTAAATATACATATAAAATGAGGACACGAAATTTTTTGGTCCCGACATTTAAAT	193753
Qy	1043	AAGTGAATATCACATATAAGAGTACACACAGTCTTTTGGCTTCCCACTGATATATAA	1102
Db	193812	AAGTAAATATACATATAAAATGAGGACACGAAATTTTTTGGTCCCGACATTTAAAT	193753
Qy	1103	GTTTTGCTTATACACACTGAGTGTGTAAAGTGTGCAATAGTATGTATGCTAAAAAA-A	1161
Db	193752	GTTATGTTTACCTTCTACTATCATCTGTGCAGGTGTGCAATACATTTATCTTAAAAATG	193693
Qy	1162	CACATACCTTATTTTAAATAGCTTTATTTACT--AAAAATGCTAACAATCATTTTGAGCA	1219
Db	193692	TACATACCTTATTTGAAAAATACATATCTCCCTAAAAAATGTAACATATCTGAGGC	193633
Qy	1220	TTCAATGATGTTGTAATCTTTTGTCTGTGTGAAGGCTTTTCTTATGATGACTGAT----	1275
Db	193632	TGCAAGGAGTCATATCTTTTGTCTGTGTACAGGCTCTTGCCCTCACTGAGATGGCTGCTG	193573
Qy	1216	-----CGGGGCTCAGGTCGCTGAAGCTTAAAGGTGGCTGTGCGACTTTCTTAA-----A	1323
Db	193572	ATTGATCAGGGGTGGGTTGTCTGAAGGTCTGGGATGCTGTACCAATTTCTTAAATATAGA	193513
Qy	1324	CAACGATGAATGTCATATTCAGTTGACTCTTCTCTT-TCATGAAGAATTTCTCTCTACT	1382
Db	193512	CAACACGAAGTTTCTGTCGATCGATTTGGCTCTCTTTATAGCAATGATTTCTTTGTAGC	193453
Qy	1383	GTTGTATGCTTTTGTATGACATTTTATGACACAGTAGAACTTCTTGAATAATTGA-TCNA	1441
Db	193452	ATGTGATGCTGTGTTGATGACATTTTA--CTCAGTAGAACTTCTTCAAAATTTGATGTCNA	193395
Qy	1442	TCCTCTCAACCCCTGCTCTGCTTTTAAACAACCTAAGTTAATATATATTTCTGAATCATTTG	1501
Db	193394	TCTTCTCAACCCCTGATGCTAATTTATCAACATTAAGTTATGAATAATTTCTAATATCTTTTG	193345
Qy	1502	TTTGTCATTTCAACAATTTTCAACAGTGTCTTCAACCAGAGATGATTCATCTCATTTCCGTG	1561
Db	193334	TTTGTCATTTCAACAATTTTCAACAGATCATCAACAGATGATGATCTCATATGAACC	193275
Qy	1562	AGATGGAATCTTTGGCTCATTCATAGAAGAAATTCCTCATCTGTTCAAGTTTATCATGA	1621
Db	193274	AC-----AATCTTTGGCTATCATCAATAGCAGCAACTGTGATCCGGCTGAAGTTTACCTTGA	193219
Qy	1622	GATTGACGACATTAACGATGATGCTTCCAGGCTCACTTACTTATTTAATTCAGATTCTCTTG	1681
Db	193218	GATTACAGCTATTCAGTACACCTT-----CTGACTCCAACTTCATTCCTGGATTCCTCAGT	193164



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5419.54 Seconds  
(without alignments)  
17274.699 Million cell updates/sec

Title: US-09-728-552A-3\_COPY\_19000\_21000

Perfect score: 2001

Sequence: 1 aatgcttgattccaagctc.....caggagcatcgggtcagatgc 2001

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

EST:

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_est3:\*

4: gb\_est4:\*

5: gb\_est5:\*

6: gb\_est6:\*

7: gb\_est7:\*

8: gb\_est8:\*

9: gb\_est9:\*

10: gb\_est10:\*

11: gb\_est11:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	207.6	10.4	415	1	AI609972 tct78q10.x
C 2	206.4	10.3	490	9	AQ426532 CITB1-El-
C 3	205.2	10.3	463	9	AQ35344 RPT-11-3
C 4	204.2	10.2	412	2	BE062478 QV4-BT025
C 5	202	10.1	590	9	AQ421062 RPT-11-1
C 6	201.6	10.1	390	5	BU588888 AGENCOURT
C 7	200.8	10.0	711	11	CR959688 Homo sapi
C 8	200	10.0	533	7	CN274852 170006000
C 9	200	10.0	687	10	AG085195 Pan tlog1
C 10	199.6	10.0	660	9	AQ376970 RPT-11-15
C 11	199.2	10.0	655	9	AQ112243 CIT-HSP-2
C 12	199.2	10.0	655	9	CA439825 UI-H-D10
C 13	199.2	10.0	655	9	CA439825 UI-H-D10
C 14	198.8	9.9	467	5	BU735183 UI-E-DW0
C 15	198.8	9.9	439	1	AQ469410 CITB1-El-
C 16	198.6	9.9	439	1	AA515728 ng70f04.s
C 17	198.4	9.9	455	1	AA904211 od88e02.s
C 18	198.4	9.9	521	1	AA904211 od88e02.s
C 19	198.4	9.9	616	9	AO554309 RPT-11-4
C 20	198.4	9.9	990	3	BM803650 AGENCOURT
C 21	198.2	9.9	711	9	AQ415030 RPT-11-2
C 22	198	9.9	441	1	AA169245 zp19d03.s

C 96	193.4	9.7	575	9	AQ587429	CITBI-El-
C 97	193.4	9.7	638	6	CD673070	Fg20b08.y
C 98	193.4	9.7	677	10	AG094571	Pan trogl
C 99	193.4	9.7	691	9	AQ899804	HS 2013 A
C 100	193.4	9.7	765	6	CD516602	CD516602 AGENCOURT

## ALIGNMENTS

RESULT 1	
AI609972/c	
LOCUS	
DEFINITION	AI609972 415 bp mRNA linear EST 21-APR-1999
	ct78c10.x1 NCI CGAP HSC3 Homo sapiens cDNA clone IMAGE:2246698 3'
	similar to contains Alu repetitive element;contains element MENS
	repetitive element ;, mRNA sequence.

ACCESSION	AF060972	
VERSION	AF060972.1	GI:4619139
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 415)	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	Unpublished (1997)	Contact: Robert Strausberg, Ph.D.

**JOURNAL COMMENT**  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CNNA Library Preparation: David B. Kitzman, Ph.D.  
CNNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLND at:  
[www-bio.lnl.l.gov/bbtp/image.html](http://www-bio.lnl.l.gov/bbtp/image.html)  
Seq primer: -40UP from Gibco.

## FEATURES

FEATURES	Location/Qualifiers
source	1. .415

## ORIGIN

Query Match	10.4%	Score 207.6;	DB 1,	Length 415;
Best local Similarity	84.8%	Pred. No. 1.9e-24;		
Matches 280; Conservative	0;	Mismatches 44;	Indels 6;	Gaps 4;

QY 1216 TCTTTTATTTTTATTTTTTGAAACAGAGTCCTACTTTGTCCACCAGGCTGGAGTACA 1275  
| | | | |  
Db 400 TTTTTTTTTTTTTTTTTTTTGAAGACAAGTCGTGCTTTTCATCCAGGGCTGGAGTACA 341  
| | | | |  
QY 1276 GTGGCTGTGTCTCGGCTCACTGCACAACCTCTCCTCCAGATTCAAGCATTTCTCTG-- 1333  
| | | | |  
Db 340 GTGG-TGTATCTCGGCTCACTGCACAACCTC-GCTTCCAGTTTCAAAGCATTTCTCTGCC 283  
| | | | |  
QY 1334 TAGACTCCGAATAGCTGGATTACAGGCGCATGCACACA--TGCTTAATTTTTGTA 1391  
| | | | |

Db	282	TCAGCTTCCCAAGTAGCTGGGATTACAGCGCGCATGCCACACACC	CGGCTAATTTTGTGA	223
Qy	1392	TTTTTAGTAGAGACAGAGTTTCGCGCATGTGACACAGGCTTC	CTTTGAACTCTCTGACTTCA	1451
Db	222	TTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGA	ACTCCTGACTTCA	163
Qy	1452	GGTGATCCACCCACTCAGCCTCCCAAGCACTGGGATTACAGCAT	TAGCACAACCTGTCC	1512
Db	162	GGTATCCGCCACACTCGGCTCCCAAGTCTGGGATTACAGGCA	MACCATCATCTCC	103
Qy	1512	CAGCCTGTTTCTCAGATCCGTGATTTGTT	1541	
Db	102	CAGCTCTCTTTCTTATATCAACTTCTT	73	

## RESULT 2

LOCUS	490 bp	DNA	linear
AC026532		GSS 24-MAR-1995	
DEFINITION	CITBI-E1-2574H2 TR CITBI-E1 Homo sapiens genomic clone 2574H2, genomic survey sequence.		

ACCESSION	A0426532
VERSION	A0426532.1
KEYWORDS	GI:4499300
SOURCE	GSS.
ORGANISM	Homo sapiens (human)

REFERENCE  
AUTHORS  
1 (bases 1 to 490)  
Zhao, S., Adams, M.D., Niernan, W., Malek, J., Shizuya, H., Simon, M. and

TITLE	Use of BAC End Sequences from Caltech Libraries for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Contact: Shaying Zhao, William Niernman, Mark Adams

COMMENT Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850

## FEATURES

FEATURES	Location/Qualifiers
source	1. 490

ORIGIN

Query Match	10.3%	Score 205.4	DB 9	Length 490
Best Local Similarity	78.0%	Pred. No. 2.9e-24		
Matches 287; Conservative	0	Mismatches 76	Indels 5	Gaps 3

QY	1155	AAAGAGAGCCATTTCA	GTGGAAATGTT	CAGAAAGTATTTATGTTTCTCAG	1214
Db	433	AAGGAAATCCATTTCC	AGATTAAAGTCA	TTTAAAGAAAAACAATTAATGCAAGTGC	374
QY	1245	ATCTTTTATTTTATTT	TTTTTGAACAGACTCT	CACTTGTCCACCAAGCTGGAGTAC	1274
Db	373	GTTATTTATTTATTT	ATTTATTTAGACAGAGCT	CACTCTGTCCACCAAGGCTGGAGTGC	314
QY	1275	AGTGGCTGTGCTCGG	CTCACTGCAACTCTGCT	CTCCAGGTTAAAGGATTTTCTCG-	1333



Db 313 TGTGAC-ACGATCTGGCTCAGTCAGCACTCCGCTCCCGGTTGAAGCATTTCTTCG 255  
Qy 1334 -TCAGTTCGCCGATAGCTGGGATTAACAGCCGATC--ACCAACATGCTTAATTTTGT 1390  
Db 254 CTCAGCTTCCGAGTGTCTGGATTAACAGCGCCGCCACACCCCACTAATTTTGT 195  
Qy 1391 ATTTTATAGAGACAGAGTTTTCGACATTTTGAACGAGCTTGCTTGAATCTCTGACTTC 1450  
Db 1394 ATTTTATAGAGACAGCGCTTTCACCATTTGACGAGCTGTCTGAACTCTGACCTC 135  
Qy 1451 AGTGATCCAGCCACTCAGCTCCCAAGCACTGGATTAACGAGCATGACCAAGCTGC 1510  
Db 134 AGTGATCCAGCCACTCAGCTCCCAAGCTGTGGATTAACGAGCATGACCAAGC 75  
Qy 1511 CCAGCTTG 1518  
Db 74 TCGGCTG 67

RESULT 3  
A0535344 463 bp DNA linear GSS 18-MAY-1999  
LOCUS RPCI-11-356115.TV RPCI-11 Homo sapiens genomic clone  
DEFINITION RPCI-11-356115, genomic survey sequence.  
ACCESSION A0535344  
VERSION A0535344.1 GI:4847034  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 463)  
Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and  
Venter,J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
Other GSSs: RPCI-11-356115.TJ  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pierre@jeong.med.buflalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buflalo.edu/ordering>) or from  
Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page:  
[http://www.tigr.org/cdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: T7  
Class: BAC ends.

FEATURES  
Source  
Location/Qualifiers  
1..463  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7636526"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-356115"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_1ib="RPCI-11"  
/note="Vector: PBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPCI11 Human Male BAC Library"

ORIGIN  
Query Match 10.3%; Score 205.2; DB 9; Length 463;  
Best Local Similarity 73.5%; Pred. No. 4.7e-24;  
Matches 302; Conservative 0; Mismatches 104; Indels 5; Gaps 3;  
Qy 1173 TGTGAAATTGTTCAAGAGATTTGATTGTTCTTCAGATCTTTTATTATT 1232

Db 33 TGTGTTATATTTTCAGACTCTCTATTTGTATCTATGACTGTCCATCTTTT 92  
Qy 1233 TTTTGAACAGAGTCTCACTTTGTACCCAGGCTGAGATGAGTGCTGTGCGC 1292  
Db 93 TTTTGAAGAGAGTCTCTCTGTGTTGCTCAGGCTGAGTGAAGTGC-GTGATTTCAAGC 151  
Qy 1293 TCACGTCAACCTTCCTCCCTCCAGGTTCAAGGATCTCTCT--TCAGCTTCCGAAATAC 1350  
Db 152 TTACTGAACCTTCCTCCCTCCAGGTTCAAGGATCTCTCTCCAGCTCCCAAGTGC 211  
Qy 1351 TGGATTAACAG--GGCATGACCAACAGGCTTAATTTTATTTTATAGAGACAGA 1408  
Db 212 TGGATTAACAGACACATGACATGCCCCGACTAATTTTGTATTTTATAGAGACAG 271  
Qy 1409 GTTTCGCAATGTTACACAGGCTTGTCTTGAATCTCTGACTTCAAGTATGACCA 1468  
Db 272 GTTTCACCATGTTGCGCAGGCTGTCTTGAATCTCTGACTTCAAGTATGACCA 331  
Qy 1469 AGCTTCCCAAGACTGGGATTAACAGCATGAGCCAGCTGCGCAAGCTTTTTCAGA 1528  
Db 332 GGCCTCCCAAGAGTGTGGATTAACAGCATGAGCACTGCGCTGCGCACTGTTCATCMT 391  
Qy 1529 TCGTATTTTGTCTGAAGCTTCAATTTCTTCTTATTTATTTTGA 1579  
Db 392 TTTTGTGTTGTTGTTTGAATGAGATTTTCATCTTGTGCGCAAGCTGGA 442

RESULT 4  
BE062478/c 412 bp mRNA linear EST 09-JUN-2000  
LOCUS BE062478  
DEFINITION QV4-BT0257-261099-011-b01 BT0257 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BE062478  
VERSION BE062478.1 GI:8407128  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 412)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=bt2=QV4-BT0257-261>)  
099-011-b01ct3=1999-10-26ct4=1  
Seq primer: puc 18 forward  
High quality sequence start: 6  
High quality sequence stop: 412.

FEATURES  
Source  
Location/Qualifiers  
1..412  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_1ib="BT0257"

/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 10.1%; Score 204.2; DB 2; Length 412;  
Best Local Similarity 81.3%; Pred. No. 76-24; Indels 5; Gaps 3;  
Matches 274; Conservative 0; Mismatches 58; Indels 5; Gaps 3;

1189 AGAAGTATTGATGATGTTTCTCAGATCTTTTATTTTATTTTGAAGAGAGTC 1248  
DB AGTACTTATATCACTGTTTCTTTCTTTCTTTTCTTTTGAAGAGAGTC 342  
1249 TCACCTTGTACCCAGGCTGAGTACAGTGGCTGTGCTGGCTCAGTCAACTCTGC 1308  
DB TCACCTGTGTGCTGAGGCTGAGTACAGTGGC-ATGCTCTCGCTCAGTCAACTCTGC 283  
1309 CTCCAGGTTCAAGGATTTCTCTG--TCAGCTCCCGAATAGTGGATTCAGAGCGCA 1366  
DB CTCCCGGTTCAAGGATTTCTCTGCTCAGCTCTCTGAGTACTGGAGTCAAGCGCTG 223  
1367 TGC--ACCAACATGCTATTTTGTATTTTGTAGTACAGAGTTCGCGATGTGAC 1424  
DB TGCCACCATGCGCAGCTATTTTGTATTTTGTAGTACAGAGGTTTACACAGCTGAGC 163  
1425 CAGGCTTGTCTGAACTCTGACTTCAAGTATTCACCACTCAGCTCCCAAGCACT 1484  
DB CAGGCTGTGTGAACTCTGACTTCAAGTATTCACCACTCAGCTCCCAAGAGTGTCT 103  
1485 GGGATTACAGGATGAGCCAGCTGCGCCAGCTGTT 1521  
DB GGGATTACAGGATGAGCCAGCTGCGCCAGCTGTT 66

RESULT 5 590 bp DNA linear GSS 23-MAR-1999  
LOCUS AQ421062  
DEFINITION RPCI-11-188010.TV RPCI-11 Homo sapiens genomic clone  
ACCESSION AQ421062  
VERSION AQ421062  
KEYWORDS AQ421062.1 GI:4478786  
SOURCE GSS.  
ORGANISM Homo sapiens (human)  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 590)  
Zhang, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P. and  
Venier, J.C.  
Venter, J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
Other GSSes: RPCI-11-188010.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Bukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeet1gr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
pdejong@jgi.doe.gov, pdejong@jgi.doe.gov, pdejong@jgi.doe.gov  
BACAC Resources (http://bacpac.med.jhu.edu/ordering) or from  
Research Genet cs (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: Sp6  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1..590  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7572153"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-188010"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_id="RPCI-11"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC library"

## ORIGIN

Query Match 10.1%; Score 202; DB 9; Length 590;  
Best Local Similarity 84.1%; Pred. No. 1.5e-23;  
Matches 264; Conservative 0; Mismatches 45; Indels 5; Gaps 3;

1207 TTTCTCAGATCTTTTATTTTATTTTATTTTGAAGAGAGTCTCACTTGTACCCAGGC 1266  
DB TGTATTACTTTCTTTCTTTCTTTTATTTTGTAGAGACAGAGTCTGCTGTGCGCCAGGC 205  
1267 TGGAGTACAGTGGCTGTGCTCGGCTCAGTCAACCTCTGCTCCGAGTTCAAGCAT 1326  
DB TGAAGTACAGTGGC-GCATCTCGGCTCAGTCAACCTCTCGGCTTCAAGCAT 264  
1327 TCTCCTG--TCAGCTCCCGAATAGTGGATTCAGAGCGATGACACCATGCG-CTA 1382  
DB TCTCCTGCTCAGCTCTCTGAGTACAGGATTCAGAGCGATGACACCATGCGCTA 324  
1383 ATTTTGTATTTTGTAGTACAGAGTTCGCACTGTGACAGGCTTGTGTAATCT 1442  
DB ATTTTGTATTTTGTAGTACAGAGTTCGCACTGTGAGTCAAGCTGTGCTTGTGAATCT 384  
1443 CTGACTTACAGTATTCACCACTCAGCTCCCAAGAGTGGATTCAGAGCATGAGC 1502  
DB CTGACTTACAGTATTCACCACTCAGCTCCCAAGAGTGGATTCAGAGCATGAGC 444  
1503 CACCGTGGCCAGCC 1516  
DB CACCGTGGCCAGCC 458

RESULT 6 390 bp mRNA linear EST 20-SEP-2002  
LOCUS BUS88888/c  
DEFINITION AGENCOURT\_8910326 NIH\_MGC\_141 Homo sapiens cDNA clone IMAGE:6387393  
ACCESSION BUS88888  
VERSION BUS88888  
KEYWORDS BUS88888.1 GI:23240250  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 390)  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: NCI  
CDNA Library Preparation: Michael Brownstein Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LCM2587 row: c column: 10  
High quality sequence scop: 389.  
Location/Qualifiers  
1..390

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6387393"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="MDH08 (T1-phage-resistant)"
/clone_lib="NIH_MGC_141"
;
/note="Vector: pNR-Lit; Site:1:StiI (ggccattagggc);
Site 2:StiI (ggcgctcgccg); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
4.4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
used in cloning as follows:
5'-AAGCAGTGGTATCATACCGACGACGATGGCCATTAGGCGCGG-3' and
5'-ATTCTAGAGCGCGGCGGCGGCGGACATG-dT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.2-0.5
kb size fraction (other fractions present in NIH MGC 142).
Library created in the laboratory of M. Brownstein (NIH,
NIH). Note: this is a NIH_MGC Library."

```

ORIGIN		
Query Match	10.1%;	Score 201.6; DB 5; length 390;
Best Local Similarity	77.2%;	Pred. No. 1.9e-23;
Matches 284; Conservative	0;	Mismatches 79; Indels 5; Gaps 3

RESULT 7	
CR959688	
LOCUS	CR959688
DEFINITION	CR959688 711 bp DNA linear GSS 10-JUN-2005
ACCESSION	Homo sapiens BAC end sequence of RZPB8737H082009D from genomic
VERSION	library (orig. Pieter J. de Jong library RPI1-11), genomic survey
KEYWORDS	sequence.
SOURCE	CR959688
ORGANISM	CR959688.1 GI:66957175
	GSS.
	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
	Homnidae; Homo.
REFERENCE	1 (bases 1 to 711)

AUTHORS	Schwarz, F., Neubert, P., Schneider, D., Peters, M. and Korn, B.
TITLE	Direct Submission
JOURNAL	Submitted (07-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer

## ORIGIN

Query Match	10.0%;	Score 200.8;	DB 11;	Length 711;
Best Local Similarity	67.3%;	Pred. No. 2.2e-23;		
Matches 327; Conservative	0;	Mismatches 154;	Indels 5;	Gaps 3;

LOCUS	533 bp	mRNA	linear	EST 16-MAY-2004
RESULT 8				
CN274852				
DEFINITION	17000600044196 GRN_PRENUE Homo sapiens cDNA 5', mRNA sequence.			
ACCESSION	CN274852			
VERSION	CN274852.1	GI:47291266		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Bradenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalian, R., Lebkowksi, J. and Stanton, L.W.			
AUTHORS	Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation			
TITLE	Nat. Biotechnol. 22 (6), 707-716 (2004)			
JOURNAL	1 (bases 1 to 533)			
PUBMED	Contact: Bradenberger R			
COMMENT	Regenerative Medicine			
	Genon Corporation			
	230 Constitution Drive, Menlo Park, CA 94025, USA			
	Tel: 650 473 8658			
	Fax: 650 473 7760			
	Email: rbradenberger@genon.com			
FEATURES	Insert Length: 533 Std Error: 0.00.			
source	Location/Qualifiers			
	1..533			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/tissue_type="Embryonic stem cell, retinoic acid and			
	mitogen-treated hES cell line H7"			
	/clone_lib="GRN PRENU"			
	/note="Oligo dt primed, full-length enriched cDNA library			
	from hES cell line H7 (p29) maintained in feeder-free			
	conditions. Embryoid bodies were generated in the presence			
	of all-trans retinoic acid and mitogens."			
ORIGIN				
Query Match	10.0%; Score 200; DB 7; Length 533;			
Best Local Similarity	80.7%; Pred. No. 3.3e-23;			
Matches	271; Conservative 0; Mismatches 60; Indels 5; Gaps 3;			
1190	GAAGTATTGATTATGTTTCTCAGATCTTTTATTTTATTTTGAACAGAGTCT			
10	GAGTTTTTAAGTTATTATTTTAAATTAATTTATTTATTTATTTGAGACAGAGTCT			
1250	CACTTTTCACCCAGAGCTGAGTAAGAGGCGTGTCTCGGCTACAGCAACCTCTCC			
70	CACCTGTGCGCCAGAGCTGAGTAAGAGG- TGATATCTTGGCTACAGCAACCTCGCC			
1310	TCCAGAGTTCAAGCATCTCTCTG-TCAGCTTCCGGAATAGCTGGATTACAGCGCAT			
129	CCCCAGTTCAAGCAATCTTGTGCTGAGCTCCCTAGTAGCTGGAATTACAGCAACT			
1368	GCACCACCA- TGCCTAATTTTGTATTATTTAGTAGAGACAGAGTTTCCCATGTTGAC			
189	GCACTGACACCTGGCTAATTTTGTATTATTTTGTAGTAGACAGAGTTTCCCATGTTGAC			
1426	AGCGTTGCTTGAATCTCTGATCTTCAAGTGATCCACCACCTCAGCTCCCAAAGACTG			
249	AGCGTGCTTGTAGCTCTGACCTCAGGATGACCAACCACTCGGCTCCCAAAGTCTG			
1486	GGATTACAGAGCATGAGCCACCGTCCCAAGCTGTTT 1521			
309	GGATTACAGGTGTGAGCCACCATGCTGAGCAAGTT 344			

LOCUS	AG085195/c	687 bp	DNA	linear	GENS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-083C18.F, genomic survey sequence.				
ACCESSION	AG085195				
VERSION	AG085195.1	GI:16636997			
KEYWORDS	GSS.				
SOURCE	Pan troglodytes (chimpanzee)				
ORGANISM	Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominae; Pan.				
REFERENCE	1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
AUTHORS	BAC end sequences of Library PTB				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 687)				
REFERENCE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.				
AUTHORS	Direct Submission				
TITLE	Submitted (02-AUG-2001) Aeo Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimbes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)				
JOURNAL	Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.				
COMMENT	PRIMERS				
	Sequencing: -21M13				
	LIBRARY				
	Vector : pKS145				
	R.Site 1 : SacI				
	R.Site 2 : SacI.				
FEATURES	location/Qualifiers				
source	1..687				
	/organism="Pan troglodytes"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9598"				
	/clone="PTB-083C18.F"				
	/sex="male"				
	/cell_type="lymphoblast"				
	/clone_lib="PTB Chimpanzee Male BAC Library"				
ORIGIN					
	Query Match 10.0%; Score 200; DB 10; Length 687;				
	Best Local Similarity 84.0%; Pred. No. 3,1e-23;				
	Matches 262; Conservative 0; Mismatches 45; Indels 5; Gaps 3;				
OY	1215 ATCTTTTAAATTTTATTTTGAACAGAGCTCACTTTGTACACCAAGCTGGAGTAC 1274				
DB	451 ATTTATTTTACTGTTTTTTTGAAGCAGAGTCTCACTGTCAACCAAGCTGGAATGC 392				
OY	1275 AGTGGCTGTGCTCGGCTCACTGCAACCTGTGCTCCAGGTTCAAGCAATTCCTCG- 1333				
DB	391 AGTGGC-CTGATCTCGGCTCACTGCAACCTGTGCTCCAGGTTCAAGCAATTCCTCG 333				
OY	1334 -TCAGCTTCCGGAATGCTGGGATTAACAGCGCATGACCAACATGC--CTAATTTTGT 1390				
DB	332 CTCGGCTCTTGAAGTGGGACTACAGACGATGCCACACACCAAGCTAATTTTGT 273				
OY	1391 ATTTTATAGACAGAGTTTGCATGTTGACCAAGCTTGACCTTGAATCTGATTC 1456				
DB	272 ATTTTATATAGACAGAGGTTTCAACATGTTGGCAAGCTGTCTCAAACTCTGACCTC 213				
OY	1451 AGTGATCACCACTCAAGCTCCCAAGACACTGGGATTAACAGCATGAGCCACCGTC 1510				
DB	212 AGTGATCACCACTTGGGCTCCCAAGTGTGGGATTAACAGCGTGAACACACAC 153				
OY	1511 CCAAGCTGTTT 1522				
DB	152 CCGGCTAATTT 141				

RESULT 10  
AQ376970/c 660 bp DNA linear GSS 20-MAY-1999  
LOCUS  
DEFINITION  
RPC111-151K23.TV RPC1-11 Homo sapiens genomic clone RPC1-11-151K23,  
genomic survey sequence.  
ACCESSION  
AQ376970  
VERSION  
AQ376970.1 GI:4347993  
KEYWORDS  
GSS.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE  
1 (bases 1 to 660)  
Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and -  
Venter,J.C.  
Use of BAC End Sequences from Library RPC1-11 for Sequence-Ready  
Map Building  
JOURNAL  
Unpublished (1997)  
COMMENT  
Other\_GSSs: RPC111-151K23.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPC1-11. For BAC  
library availability, please contact Pieter de Jong  
(piteredejong.med.bufileo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: SP6  
Classes: BAC ends.  
FEATURES  
Location/Qualifiers  
source  
1..660  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7557862"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-151K23"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_1fb="RPC1-11"  
/note="Vector: pBACe3.6; Site 1: EcoRI; site\_2: EcoRI;  
RPC111 Human Male BAC Library"

ORIGIN  
Query Match 10.0%; Score 199.6; DB 9; Length 660;  
Best Local Similarity 80.8%; Pred. No. 3.6e-23;  
Matches 270; Conservative 0; Mismatches 59; Indels 5; Gaps 3;  
OY 1194 TTTTGGATATGTTTTCAGATCTTTTATTTTATTTTGAACAGAGCTCACT 1253  
DB 372 TATTTTATAAAGATTTCATATACTTTTTTTTTTTTTTTTGGATGGAGTTCACAT 313  
OY 1254 TTGTCAACCCAGGCTGAGTACAGTGTGCTCGGCTCACTGCAACCTCGCTCC 1313  
DB 312 CATATCAACGAGCTGAGGAGGCAATGG-TGCATCTCGGCTCACTGCAACCTCGGCTCC 254  
OY 1314 AGGTTCAAGCAATCTCTCTG--TCAGCTCCCGAATAGCTGGGATTAACAGCGCATGCAC 1371  
DB 253 GGGTTCAGATGATCTCTCGCTCAAGCTCCCAAGTAGCTGGGATTAGAGGACATGCCA 194  
OY 1372 CACCA--TGCCTAATTTTGTATTTTGAAGACAGAGTTCCGCGCATGTGACCGAGC 1429  
DB 193 CGGCACTGCGCTATTTTGTATTTTGAAGAGGAGGAGGCTTCCGCGATTTAGCCAGGC 134  
OY 1430 TTGCCTTGAATCTGACTTCAAGTATCACCACCTGAGCTCCCAAGCACTGGAGT 1489  
DB 133 TGGTTTGAATCTCTGACGTCAAGGTATCGGCGGCTCGGCTCCCAAGTGTCTGGAT 74

OY 1490 TACAGCATGAGCACCGTCCAGCTGTTTC 1523  
DB 73 TACAGGGGTGAACCACTGTGCTCGGCTGATTTTC 40  
RESULT 11  
AQ112243 461 bp DNA linear GSS 29-AUG-1998  
LOCUS  
DEFINITION  
CIT-HSP-2373C2.TR CIT-HSP Homo sapiens genomic clone 2373C2,  
genomic survey sequence.  
ACCESSION  
AQ112243  
VERSION  
AQ112243.1 GI:3484403  
KEYWORDS  
GSS.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE  
1 (bases 1 to 461)  
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and  
Venter,J.C.  
Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building  
JOURNAL  
Unpublished (1998)  
COMMENT  
Other\_GSSs: CIT-HSP-2373C2.TF  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Classes: BAC ends.  
FEATURES  
Location/Qualifiers  
source  
1..461  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="2373C2"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_1fb="CIT-HSP"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

ORIGIN  
Query Match 10.0%; Score 199.4; DB 9; Length 461;  
Best Local Similarity 74.8%; Pred. No. 4.3e-23;  
Matches 303; Conservative 0; Mismatches 96; Indels 6; Gaps 4;  
OY 1152 ACAAAAGAGAGCCATTGCTGTAATTTTCAAGAGATTTGATATGTTTCT 1211  
DB 11 ACTCAAGCTTATCAATTTTATTTGATCACTGTTCATATGATTTCTTTCTTTCTTCT 70  
OY 1212 CAGATCTTTTATTTTATTTTATTTT--GAAAGAGCTCACTTTGTACCCAGGCTGGA 1270  
DB 71 TTTTCTTTTTTTTTTTTTTTTTTTTTTTCAGACAGAGTCTCACTGTGCTGCCCGGCGGA 130  
OY 1271 GTACAGTGGCTGTGCTCGGCTCACTGCAACCTGTGCTCCAGAGTTCAAGATTCCTC 1330  
DB 131 GAGAGGTGG-TGCATCTTGGCTCACTGCAACCTCGGCTCTAGGTACAAAGCATTCCTC 189  
OY 1331 CTG--TCAGCTTCCGAATAGCTGGGATTTACAGGCGCATGC--ACCACTAGCTTAATTT 1386  
DB 190 CTGCTTCAAGCTCCCAAGTAGCTGAGGTTACAGGCTCGCCACCATGCTCACTAATTT 249  
OY 1387 TTGTAATTTTATAGACAGAGTTTCCGCGATGTGACCAAGCTTGCTTGAATCTCTGA 1446

Db 250 TTGATTTTGTAGAAACGGGGTTTCAACATGTTGCGCAGGCTGCTTGAACCTTGA 309  
Qy 1447 CTTAGGATGATCCACCACTCAGCTCCCAAGACATGAGATTACAGCAATGAGCCACC 1506  
Db 310 CCTCAGGATGATCGGCTCGGACCTCCCAAGGATGATACAGTGTGAGCACT 369  
Qy 1507 GTGCCAGCCTGTTTCTCAGATCCTGATTTGTTTCTGAAGCCT 1551  
Db 370 GTACCCGCGCTTTATAGTATTTCTTTATTTCTTTTAAGCT 414

RESULT 12  
CA439825 655 bp mRNA linear EST 08-NOV-2002  
LOCUS UI-H-D10-auz-j-14-0-UI 81 NCI CGAP D10 Homo sapiens cDNA clone  
DEFINITION UI-H-D10-auz-j-14-0-UI 3', mRNA sequence.  
ACCESSION CA439825  
VERSION CA439825.1 GI:24804245  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 655)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mai.nih.gov  
Tissue Procurement: Dr. Jose Mercuende  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
The following repetitive elements were found in this cDNA  
sequence: 1-22, >AT\_richlow\_complexity (matched complement)  
86-110, >POLY\_A\$Simple\_repeat (matched complement) 111-400, >ALU  
(matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
SOURCE location/Qualifiers

1..655  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-D10-auz-j-14-0-UI"  
/tissue\_type="lung Focal Fibrosis"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP D10"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
NCI CGAP D10 is a cDNA library containing the following  
tissue(s): A pool of lung Focal Fibrosis. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pT73-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(ATT)18 tail. The sequence tag for this library is  
ATACGGGCTC.  
TAG\_TISSUE=lung with fibrosis  
TAG\_LIB=UI-H-D10  
TAG\_SEQ=ATACGGGCTC"

ORIGIN

Query Match 10.0%; Score 199.2; DB 6; Length 655;  
Best Local Similarity 82.1%; Pred. No. 4.2e-23;  
Matches 266; Conservative 0; Mismatches 53; Indels 5; Gaps 3;  
Qy 1197 TTGATATGTTTTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTG 1256  
Db 78 TAGAGCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 137  
Qy 1257 TCACCAAGCTGAGATACAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1316  
Db 138 TCACCAAGCTGAGATACAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 196  
Qy 1317 TTCAAGCATTTCTCTG--TAGCTTCCCAATAGCTGGATTACAG--GGCATGACCC 1372  
Db 197 TTCAAGCATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256  
Qy 1373 ACCATGCTATTTTGTATTTTGTATGAGACAGATTTGCGCATGTGTGACAGGCTTG 1432  
Db 257 GCGCTGCTAATTTTGTATTTTGTATGAGATGAGATGAGATGAGATGAGATGAGATGAG 316  
Qy 1433 CTTGACCTCTGACTGATGATCCACCACTCAGCTTCCCAAGCATGGGATTAC 1492  
Db 317 TCTGCACTCTGCGCTCAAGTATCCACCACTCAGCTTCCCAAGCATGGGATTAT 376  
Qy 1493 AGGCATGAGCCACCGTCCCAAGCC 1516  
Db 377 AGGCGTGCGCACCTGTGCGCGGCC 400

RESULT 13  
BZ603010 694 bp DNA linear GSS 08-JUN-2003  
LOCUS WHAAX35TF Human MCF7 breast cancer cell line library (MCF7\_1) Homo  
DEFINITION sapiens genomic clone MCF7\_1-6F22, genomic survey sequence.  
ACCESSION BZ603010 GI:31511472  
VERSION BZ603010.1 GI:31511472  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 694)  
Volk, S., Zhao, S., Chin, K., Brebner, J.H., Herndon, D.R., Tao, Q.,  
Kowbel, D., Huang, G., Lapuk, A., Xuo, W.-L., Magrane, G., de Jong, P.,  
Gray, J.W. and Collins, C.  
End-sequence profiling: Sequence-based analysis of aberrant genomes  
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)  
12788976  
COMMENT Contact: Volk SV  
Colin Collins' lab  
UCSF Comprehensive Cancer Center  
UCSF Box 0808, San Francisco, CA 94143-0808, USA  
Tel: 415 502 7066  
Fax: 415 502 5665  
Email: svolk@cc.ucsf.edu  
This clone is available from Amplicon Express  
http://www.genomex.com  
Classes: BAC ends.

FEATURES  
source location/Qualifiers

1..694  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="MCF7\_1-6F22"  
/sex="female"  
/clone\_lib="Human MCF7 breast cancer cell line library  
(MCF7\_1)"  
/note="Vector: pECBAC1; site 1: HindIII; This library was  
constructed from MCF7 breast cancer cell line by Amplicon  
Express (http://www.genomex.com) using their standard  
procedure."

ORIGIN





Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES  
source

1..591  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="2596F16"  
/sex="male"  
/cell\_type="sperm"  
/clone\_id="CIRBI-E1"  
/note="Vector: pBelobAC11, Site\_1: EcoRI, Site\_2: EcoRI;  
Caltech Human BAC Library D"

## ORIGIN

Query Match 9.9%; Score 198.8; DB 9; Length 591;  
Best Local Similarity 85.6%; Pred. No. 5e-23;  
Matches 256; Conservative 0; Mismatches 38; Indels 5; Gaps 3;

QY 1220 TTTATTTTATTTTATTTTGAACAAGTCTCACTTTGTACCCAGGCTGGAGTACAGTGG 1279  
|||||  
DB 590 TTTTATTTTATTTTATTTTGAACAAGTCTCACTTTGTACCCAGGCTGGAGTACAGTGG 531

QY 1280 CTGTGCTCGGCTCACTGCAACTCTGCTCCAGGTTCAAGCATTTCCCG--TCAG 1337  
|||||  
DB 530 C-GCATCTCGGCTCACTGCAACTCTGCTCCAGGTTCAAGCATTTCTGCTCTCAA 472

QY 1338 CTTCCTCGATAGCTGGATTTAAGCGCATGACCAACATGCTTAATTTTGTATTTTGA 1397  
|||||  
DB 471 CCTCCTAGTACTGGATTTACAGTGC--GGCTTACATGCTTAATTTTGTATTTTGA 414

QY 1398 GTAGAGACAGAGTTTGGCATGTTGACAGGCTTGGCTTAATCTCTGACTTCAGGTAT 1457  
|||||  
DB 413 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGCTCGAACTCTGACCTCAGGTGAT 354

QY 1458 CCACCCACTCTCAGCTCCCAAGACATGGGATTTACAGGATGAGCCACGTCGCCAGCC 1516  
|||||  
DB 353 CCACCCGCTTGGCTTCCCAAGTGTAGGATTTACAGGATGAGCCACGCTGCGCC 295

## RESULT 16

AA515728 439 bp mRNA linear EST 19-ANG-1997  
LOCUS ng70f04.s1 NCI CGAP Lip2 Homo sapiens cDNA clone IMAGE:940159  
DEFINITION similar to contains Alu repetitive element; contains element PTR5  
repetitive element; mRNA sequence.

ACCESSION AA515728  
VERSION AA515728.1 GI:2255328  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 439)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.

FEATURES  
source

cdna library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNI at:  
www-bio.illn.gov/bbrp/image/image.html  
Insert Length: 634 Std Error: 0.00  
Seq primer: -40m13 fwd. RT from Amersham  
High quality sequence stop: 358.  
Location/Qualifiers  
1..439  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:940159"  
/issue\_type="liposarcoma"  
/lab\_host="DH10B"  
/clone\_id="NCI CGAP Lip2"  
/note="Vector: pAMP10; mRNA made from liposarcoma, CDNA  
made by oligo-dT priming. Non-directionally cloned.  
Size-selected on agarose gel, average insert size 600 bp.  
Reference: Kitzman et al. (1996) Cancer Research  
56:5380-5383."

## ORIGIN

Query Match 9.9%; Score 198.6; DB 1; Length 439;  
Best Local Similarity 84.1%; Pred. No. 5.9e-23;  
Matches 260; Conservative 0; Mismatches 44; Indels 5; Gaps 3;

QY 1216 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTTGTACCCAGGCTGGAGTACA 1275  
|||||  
DB 58 TCTTTTCTTTTATTTTATTTTGAACAGAGTCTCACTTTGTACCCAGGCTGGAGTACA 117

QY 1276 GTGGCTGAGTCTCGGCTCACTGCAACTCTGCTCCAGGTTCAAGCATTTCTCTG-- 1333  
|||||  
DB 118 GTGGC-ACGATCTCAGCTCACTGCAACTCTGCTCCAGGTTCAAGCATTTCTCTGCG 176

QY 1334 TCAGCTTCCGAATAGCTGGGATTTACAG--GGCATGACCAACATGCTTAATTTTGTGA 1391  
|||||  
DB 177 TCAGCTTCCGAGTAGTGGGATTTACAGCTGACACCAACAGCTGCTTAATTTTGTGA 236

QY 1392 TTTTATAGTAGACAGAGTTTGGCATGTTGACAGGCTTGGCTTAATCTCTGACTTCA 1451  
|||||  
DB 237 TTTTATAGTAGACCGGGGTTTGGCATGTTGACAGGCTGCTCAAACTCTGACTTCA 296

QY 1452 GGTGATCCACCCAGCTCAGCTCCCAAGACATGGGATTTACAGGATGAGCCACGTCGC 1511  
|||||  
DB 297 AGTGATCCGCCAAGCTCAGCTCCCAAGTGTGGATTTAGGATGAGCCACCAACGC 356

## RESULT 17

AA904211 455 bp mRNA linear EST 21-APR-1998  
LOCUS o088e02.s1 NCI CGAP Br5 Homo sapiens cDNA clone IMAGE:1377338  
DEFINITION similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA904211  
VERSION AA904211.1 GI:3039334  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 455)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov

Tissue Procurement: Ian Kirsch, M.D., Kristina A. Cole, M.D.,  
Ph.D. student, Rodrigo F. Chuqui, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.

CDNA Library Preparation: David B. Krizman, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/ILN at:

www.bio.lnlnl.gov/bhrp/image/image.html

Insert length: 579 Std Error: 0.00

Seq primer: -40m13 fwd. RT From Amerham

High quality sequence stop: 421.

Location/Qualifiers

1. 455

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1377338"

/sex="female"

/tissue\_type="infiltrating ductal carcinoma"

/lab\_host="PH10B"

/clone\_lib="NCI-CCAP Br5"

/note="Organ: breast; Vector: pAMP10; mRNA made from

infiltrating ductal carcinoma, CDNA made by oligo-dr

priming. Non-directionally cloned. Size-selected on

agarose gel, average insert size 600 bp."

# ORIGIN

Query Match 9.9%; Score 198.4; DB 1; Length 455;  
Best Local Similarity 82.2%; Pred. No. 6.3e-23;  
Matches 264; Conservative 0; Mismatches 52; Indels 5; Gaps 3;

QY 1208 TTCTAGATCTTTTATTTTATTTTGAAGAGAGTCTCACTTGTGACCCAGGCT 1267  
DB 25 TTTCGGCGCGCTTCACATTTTATTTTGGAAACAGAGTCTCGGTGTCTCAACCCAGGCT 84  
QY 1268 GGAGAGACAGTGGCTGNGTCTCGGCTCACTGCAACCTCGCTCCAGGTTCAAGGATT 1327  
DB 85 GGAGTGCAGTGGC-ATAATCTTGGCTCACTGCAACCTCGCTCCAGGTTCAAGGATT 143  
QY 1328 CTCCTG--TCAGCTTCCGAGATAGCTGGATTACAGCGCATGACACCA--TGCCTTA 1383  
DB 144 CTCCTGCTCAAGCTTCCAGAGTAGCTGGGATTACAGGACACACACCACTGCTGCTTA 203  
QY 1384 TTTTGTATTTTATTAAGAGAGAGTTCGCCATGTTGACAGGCTTGCCTTGAACCTC 1443  
DB 204 TTTTGTATTTTATTAAGAGAGAGGAGGTTTCACTATGTTGGCCAGGCTGTCTCAAACTCC 263  
QY 1444 TGACTTCAAGTATCCAGCCTCAGCCTCCCAAGGACAGTGAATTACAGGATGAGGCC 1503  
DB 264 TGACCTCAGGTATCCACCCCTGGCTCCCAAGGATGAGGATTAAGAGTGTGAGCC 323  
QY 1504 ACCGTGCCAGGCTGTTTCT 1524  
DB 324 ACCGTGCCAGGCTGCTTCT 344

## RESULT 18

AM970571/c

LOCUS AM970571 521 bp mRNA linear EST 01-JUN-2000

DEFINITION EST382652 MAGR resequences, MAGK Homo sapiens CDNA, mRNA sequence.

ACCESSION AM970571

VERSION AM970571.1 GI:8160416

KEYWORDS EST.

SOURCE

ORGANISM Homo sapiens (human)

Homosapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo

1 (bases 1 to 521)

Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C.,

Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and

## TITLE

Assessment of gene expression patterns in a model of colon tumor

metastasis using a 19,200 element CDNA microarray

Unpublished (2000)

JOURNAL The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: john@igr.org

Plate: 279

Seq primer: Forward.

Location/Qualifiers

1. 521

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone\_lib="MAGR resequences, MAGK"

/note="Vector: pBluescriptSM"

## ORIGIN

Query Match 9.9%; Score 198.4; DB 1; Length 521;  
Best Local Similarity 81.4%; Pred. No. 6.1e-23;  
Matches 267; Conservative 0; Mismatches 56; Indels 5; Gaps 3;

QY 1195 ATTGATATGTTTCTCAGATCTTTTATTTTATTTTGAAGAGAGTCTCACTT 1254  
DB 504 ACTTTTATTTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 445  
QY 1255 TGTCAACCAAGCTGAGTACAGTGTGCTGTGCTCACTGCAACCTGTGCTTCCA 1314  
DB 444 TGTCCCAAGCTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 386  
QY 1315 GGTTCAGAGATTTCTCTG--TCAGCTTCCGAGATAGTGGATTAAGAGCCAGTC-A 1370  
DB 385 GGTTCAGAGATTTCTCTGCTTCCGAGATAGTGGATTAAGAGCCAGTC-A 326  
QY 1371 CCACATGCGTAATTTTGTATTTTATTAAGAGAGAGTTCGCAATGTAACAGGCT 1430  
DB 325 CAGCTCCGAGTATTTTGTATTTTATTAAGAGAGAGTTCGCAATGTAACAGGCT 266  
QY 1431 TGCCTTGAATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 1490  
DB 265 GGTCTTGAATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 206  
QY 1491 ACAGGATGAGCCACCGTCCAGCTTG 1518  
DB 205 ACAGGATGAGCCACCGTCCAGCTTG 178

## RESULT 19

AOS54309/c

LOCUS AOS54309 616 bp DNA linear GSS 28-MAY-1999

DEFINITION RPCI-11-41214.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-41214,

genomic survey sequence.

ACCESSION AOS54309

VERSION AOS54309.1 GI:4913486

KEYWORDS GSS.

SOURCE

ORGANISM Homo sapiens (human)

Homosapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 616)

Zhao, S., Adams, M. D., Nieman, W., Malek, J., de Jong, P. and

Venter, J. C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Map Building

Unpublished (1997)

JOURNAL Contact: Shaying Zhao, William Nieman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200



Eukaryōta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
COMMENT

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
1 (bases 1 to 751)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 3135 row: I column: 22  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 751.  
Location/Qualifiers  
1..751  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="plate=3135 Col=22 Row=I"  
/sex="male"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/note="Organ: sperm; Vector: pbeloBAC11; BAC Clones in  
E-Coli DH10B"

ORIGIN

Query Match 9.9%; Score 198; DB 9; Length 751;  
Best Local Similarity 85.1%; Pred. No. 6.4e-23;  
Matches 257; Conservative 0; Mismatches 40; Indels 5; Gaps 3;

1219 TTTTATTTTATTTTATTTTGAACAGAGCTTCACCTTGACCCAGGCTGAGTACAGG 1278  
189 TTTCTTTTATTTTATTTTGAACAGAGCTTCACCTTGACCCAGGCTGAGTACAGG 248  
1279 GCTGAGCTGCTGGCTCAGTCACTGCACTCTGCTCCAGGTTCAAGGATTCCTCTG 1336  
249 G-TGGATCTGGCTCAGTCACTCTGCTCCAGGTTCAAGGATTCCTCTGCTCA 307  
1337 GCTTCCCGAATAGCTGGGATTACAGGCGGATGACACCA--TGCCTAATTTTGTATT 1394  
308 GCCTCCCGAGTAGCTGGGATTAGGCACTGACACCACTGGCTAATTTCTGTATT 367  
1335 TTAGTAGAGACAGAGTTTGCATGTGACAGGCTTGCTGAACTCTGACTTCAGGT 1454  
368 TTAGTAGAGACAGGAGTTTGAACATGTGGCCAGGCTGCTGAACTCTGACTTCAGGT 427  
1455 GATCACCACCTCAGGCTCCCAAGACATGGGATTACAGGATGAGCCAGCTGCCAG 1514  
428 GATTCACCTGCTTGGCTCCCAAGATGTTGGATTACAGGATGAGCCAGCTGCCG 487  
1515 CC 1516  
488 CC 489

RESULT 24  
BU603903/c 322 bp mRNA linear EST 20-SEP-2002  
LOCUS AGENCOURT\_8932582 NIH\_MGC\_141 Homo sapiens cDNA clone IMAGE:6499491  
DEFINITION 5', mRNA Sequence.  
ACCESSION BU603903  
VERSION BU603903.1 GI:23255662  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
1 (bases 1 to 322)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: NCI  
CDNA Library Preparation: Michael Brownstein Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LULU)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LULU at:  
http://image.llnl.gov  
Plate: LUCM263 row: n column: 04  
High quality sequence stop: 321.  
Location/Qualifiers  
1..322  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6499491"  
/tissue\_type="mixed (pool of 40 RNAs)"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_lib="NIH MGC 141"  
/note="Vector: pDNR-LIB; Site 1: SfiI (ggccatcgagc);  
Site 2: SfiI (ggcgctggcg); Double-stranded cDNA was  
prepared from a pool of 40 cell line polyA+ RNAs (bladder  
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -  
4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,  
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,  
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary  
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were  
used in cloning as follows:  
5'-AAGCATGATGATCAACGAGTGGCCATTACGCGCGG-3' and  
5'-ATTCTAGAGCGCGAGCGCGCGAGCATG-dT(30)NN-3'. Full-length  
enriched library was constructed using the Clontech  
Creator SMART kit and size-selected to contain the 0.2-0.5  
kb size fraction (other fractions present in NIH MGC 142).  
Library created in the laboratory of M. Brownstein (NIMH,  
NIH). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 9.9%; Score 197.8; DB 5; Length 322;  
Best Local Similarity 84.4%; Pred. No. 8.6e-23;  
Matches 259; Conservative 0; Mismatches 42; Indels 6; Gaps 3;

1216 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCAC-TTGTGCAACCGAGCTGAGTAG 1274  
309 TTTTATTTTATTTTATTTTGAACAGAGTCTCAC-TTGTGCAACCGAGCTGAGTAG 250  
1275 AGTGCCTGCTGCTGCTCACTGCACTCTGCTCCAGGTTCAAGGATTCCTCTG 1333  
249 AATGGCGGATCTGGCTCACTGCACTCTGCTCCAGGTTCAAGGATTCCTCTG 190  
1334 -TCAGTCCCGAATAGCTGGGATTACAGGCGGATG---CAGCACATGCTAATTTTGG 1389  
189 TTCAGCTCTCGAGTAGCTGGGATTACAGGTCCTGACACACGAGCCCGGTAAATTTTGG 130  
139C TATTTTATGAGACAGAGTTTCCCATGTTGACAGGCTTGCTTGAATCTGACTT 1449  
129 TATTTTATGAGACAGGAGTTTCAACAGTGTGTCAGGCTGTCTGCAACTCTGACT 70  
1450 CAGGTGATCCACCACTTCAGGCTCCCAAGACATGGGATTACAGGATGAGCCAGCTG 1509  
69 CAGGTGATCCACCACTTCAGGCTCCCAAGAGTGTGAGATTACAGGCTGAGCCACTGTG 10  
1510 CCCAGCC 1516  
9 CCCAGCC 3

RESULT 25  
 BUS88173/c

LOCUS BUS88173 324 bp mRNA linear EST 20-SEP-2002  
 DEFINITION AGENCOURT 8952135 NIH\_MGC\_141 Homo sapiens cDNA clone IMAGE:6386321  
 5', mRNA sequence.

ACCESSION BUS88173  
 VERSION BUS88173.1 GI:23239452  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo  
 1 (bases 1 to 324)  
 NIH-MGC <http://mgc.mci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: NCI  
 CNA Library Preparation: Michael Brownstein Laboratory  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2584 row: f column: 18  
 High quality sequence stop: 323.  
 Location/Qualifiers

FEATURES  
 source 1..324  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6386321"  
 /tissue\_type="mixed (pool of 40 RNAs)"  
 /lab\_host="DH10B (Tl-phage-resistant)"  
 /clone\_lib="NIH\_MGC\_141"  
 /note="Vector: pDNR-LIB; Site\_1: SfiI (ggccattatggcc); Site\_2: SfiI (ggccgctcgcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:  
 5'-AAGCAGTGTATCAAGCAGAGTGGCATTTACGCGGG-3' and 5'-ATTCAGAGCCGAGCGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART Kit and size-selected to contain the 0.2-0.5 kb size fraction (other fractions present in NIH MGC 142). Library created in the laboratory of M. Brownstein (NIH). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 9.9%; Score 197.8; DB 5; Length 324;  
 Best Local Similarity 80.6%; Pred. No. 8; 6e-23;  
 Matches 257; Conservative 0; Mismatches 57; Indels 5; Gaps 2;

OR 1203 ATGTTTCTCAGATCTTTTATTTTATTTTGAAGAAGAGTCTCACTTGTGACCC 1262  
 DB 321 AGTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGAGTTTCACTTTGTTCACCC 262

OR 1263 AGGCTGAGTACAGTGTGTGTCTCGGCTCACTGCAACCTCTGCCAGGTTCAAG 1322  
 DB 261 AGCGTGAAGTGAAGGGCGTATCTCGGCTCACTGCAACCTCTGCCAGGTTCAAG 202

OR 1323 CGATTTCCTG--TAGCTTCCGAAATAGCTGGATTTACAGGCGCATG--CACCACAT 1377  
 DB 201 TGATTCTCTGTTTACGCTCTGAGTGTGAGTTACAGGTGCTGCAACACGAGGCC 142

OR 1378 GCGTAATTTTGTATTTTATTTAGAGACAGAGTTTGGCATTGTGACAGCTTGCTTG 1437  
 DB 141 GGTATTTTGTATTTTATTTAGTAAGAAGGGGTTTACCACTTGTGTAGGCTGTG 82

OR 1438 AACTCTGACTTCAAGGTATTCACCACTTCAAGCTTCCCAAGCACTGGGATTTACGGCA 1497  
 DB 81 AACTCTGACTTCAAGGTATTCACCACTTCAAGCTTCCCAAGCACTGGGATTTACGGCG 22

OR 1498 TGAGCCACGTCGCCAGCC 1516  
 DB 21 TGAGCCACTGTGCCAGCC 3

RESULT 26  
 A0528478/c

LOCUS A0528478 757 bp DNA linear GSS 18-MAY-1999  
 DEFINITION RPCI-11-349N11.TV RPCI-11 Homo sapiens genomic clone  
 RPCI-11-349N11, genomic survey sequence.

ACCESSION A0528478  
 VERSION A0528478.1 GI:4840591  
 KEYWORDS GSS.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo  
 1 (bases 1 to 757)  
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and  
 Venter, J.C.  
 Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
 Map Building  
 Unpublished (1997)  
 Other GSSs: RPCI-11-349N11.TU  
 CONTACT: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: [hbe@tigr.org](mailto:hbe@tigr.org)  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Piter de Jong  
 (piterdejong.med.buffalo.edu). Clones may be purchased from  
 BACDAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
 Research Genes, Inc. ([info@resgen.com](http://info@resgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html).  
 Seq primer: 77  
 Class: BAC ends.  
 Location/Qualifiers

FEATURES  
 source 1..757  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7631954"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-349N11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /clone\_lib="RPCI-11"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPCI11 Human Male BAC Library"

ORIGIN

Query Match 9.9%; Score 197.8; DB 9; Length 757;  
 Best Local Similarity 79.1%; Pred. No. 6; 9e-23;  
 Matches 273; Conservative 0; Mismatches 67; Indels 5; Gaps 3;

OR 1197 TTGATTATGTTTCTCAGATCTTTTATTTTATTTTGAAGAAGAGTCTCACTTGTG 1256  
 DB 589 TTGATTATGTTTCTTACATTTCTTTTATTTTATTTTGAAGAAGAGTCTCACTCTG 530

OR 1257 TCACCAGCTGAGTACAGTGTGTGTCTCGGCTCACTGCAACCTCTGCTCCAGG 1316  
 DB 529 TTACCAGCTGAGTGTGAGTGTG--TGCAGCTTGGCTCACTGCAACCTCTCGGCTACGGG 471

RESULT	27
LOCUS	AO738849/c
DEFINITION	AO738849 896 bp DNA linear GSS 16-JUL-1999
ACCESSION	HS_5388_B1.G06 T7A RPCR-11 Human Male BAC library Homo sapiens
VERSION	genomic clone Plate=962 Col=11 Row=N, genomic survey sequence.
KEYWORDS	AO738849.1 GI:5516371
SOURCE	GSS.
ORGANISM	Homo sapiens (human)
	Homo sapiens

REFERENCE AUTHORS	TITLE	JOURNAL PUBMED	COMMENT
1 (bases 1 to 896) Mahatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood L.	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)	
Contact: Mahatras GG, Wallace JC, Hood L		10449764	

TITLE	JOURNAL	PUBLISHED	COMMENT
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	Proc. Natl. Acad. Sci. U.S.A.	96 (17), 9739-9744 (1999)	
10449764	Contact: Mahairas GG, Wallace JC, Hood L	High Throughput Sequencing Center	
	University of Washington	401 Queen Anne Avenue North, Seattle, WA 98109, USA	
	Tel: (206) 616-3618		
	Fax: (206) 616-3887		
	Email: jwallace@u.washington.edu		
	Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACpac Resources ( <a href="http://bacpac.med.buffalo.edu/ordering_bac.htm">http://bacpac.med.buffalo.edu/ordering_bac.htm</a> ) or from Research Genetics ( <a href="http://www.resgen.com">http://www.resgen.com</a> ). BAC end web Server: <a href="http://www.htsc.washington.edu">http://www.htsc.washington.edu</a>		
	Plate: 962	row: N	column: 11
	Seq primer: 77		
	Class: BAC ends		
	High quality sequence	strop: 896.	
	Location/Qualifiers		
	1..896		

ORIGIN	
Query Match	9.9%; Score 197.8; DB 9; Length 896
Best Local Similarity	79.1%; Pred. No. 6.7e+23;

[illegible]

RESULT 28	
BX486751	
LOCUS	BX486751 493 bp mRNA linear EST 04-SEP-2003
DEFINITION	DKEZP686E18253_r1 686 (synonym: hlc3) Homo sapiens cDNA clone
ACCESSION	DKEZP686E18253_5', mRNA sequence.
VERSION	BX486751
KEYWORDS	BX486751.1 GI:31950724
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 493) Bloecker,H., Boecker,M., Mewes,H.W., Weill,B., Amid,C., Osanger,A., Robo G., Han,M. and Wiemann,S. EST (Bloecker,H., Boecker,M., Mewes,H.W., Weill,B., Amid,C., et al.) Unpublished (2003) Contact: MIPS
TITLE	
JOURNAL	
COMMENT	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 493)	Bloeker, H., Boecher, M., Mewes, H. W., Weill, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.	EST (Bloeker, H., Boecher, M., Mewes, H. W., Weill, B., Amid, C., et al.)	Unpublished (2003)	Contact: MIPS
MIPS	Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany	This is the 5' sequence of the clone insert	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;	sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the CDNA sequencing consortium of the German Genome Project.
No sl sequence available.	This clone (DKFZ666E1825) is available at the RZPD in Berlin.	Please contact the RZPD: Ressourcenzentrum, Heubergweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.		

ORIGIN	Score	DB	Length
Query Match	9.9%	197.6	5
Best Local Similarity	72.3%	Pred. No. 8.3e-23	493







DEFINITION RPCI-11-177C16.TV RPCI-11 Homo sapiens genomic clone  
RPCI-11-177C16, genomic survey sequence.  
ACCESSION AQ415537  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 690)  
Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and  
Venter,J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other\_GSSs: RPCI-11-177C16.TU  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genet. cs (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: T7  
Class: BAC ends.  
FEATURES  
source Location/Qualifiers  
1..690  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7567647"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-177C16"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPCI11 Human Male BAC Library"  
ORIGIN  
Query Match 9.9%; Score 197.4; DB 9; Length 690;  
Best Local Similarity 78.3%; Pred. No. 8.3e-23;  
Matches 275; Conservative 0; Mismatches 71; Indels 5; Gaps 3;  
QY 1179 AATTGTTTCAGAGAGATTGATTGATTGTTTCTCAGATCTTTTATTTTATTTTATTTT 1238  
DB 379 AGTGTGTAGAACAGGCTCTCAGAGAGTGTATGCTTTTGTGTTTGTGTTTGTG 320  
QY 1239 AAAAGAGTCTCAGTCTGTCAACCCAGGCTGAGATAGAGGCGCTCCGGCTACAG 1298  
DB 319 AGACAGAGTTTCACTCTGTGCTCCAGGCTGAGAGTGAAGTGG--TACATCTGTGCTAC 261  
QY 1299 CAACCTCTGCTCCAGGTTCAAGCGATTCTCCTG--TCAGCTTCCCGAATAGCTGGAT 1356  
DB 260 CAACCTCTGCTCCAGGTTCAAGCGATTCTCCTGCTCAGCTCCCGAGTAGCTGGAT 201  
QY 1357 TACAGCGCGATG--ACACCATGCTTAATTTTGTATTTTATTTTATTTTATTTTATTT 1414  
DB 200 TACAGGTGACACACACACTCCCGGCTAATTTTGTATTTTATTTTATTTTATTTTATTT 141  
QY 1415 CCATGTTGACAGGCTTGCCTTGAATCTCCGATTCAGAGTATCCACCACCTCAGCTTC 1474  
DB 140 CCATGTTGCTCAGGCTGTGCTTGAATCTCGAATCTGACCTGATCCACTTCTCAGCTTC 81  
QY 1475 CCAAGACCTGGAGATTACAGGATGAGCCACCGTGCCTGAGCTGTTTCTC 1525  
DB 80 CCAAGTGTGGGATTACAGGATGAGCCACCGCTGCTGATATTTGTC 30

RESULT 33  
CL423123/c 720 bp DNA linear GSS 12-MAR-2004  
LOCUS RPI1-429A8, T7 RPCI-11 Human Male BAC Library Homo sapiens genomic  
DEFINITION clone RPI1-429A8, genomic survey sequence.  
ACCESSION CL423123  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 720)  
Zhao,S. and Yonescu,R.  
TITLE Direct Submission of BAC End Sequence of CCAP clones  
JOURNAL Unpublished (2003)  
COMMENT Contact: Yonescu, R  
NCI/NIH  
9000 Rockville Pike, Bethesda, MD 20892, USA  
Tel: 301-402-2008  
Fax: 301-402-1204  
Email: Yonescu@pop.nci.nih.gov  
Plate: 429 row: A column: 8  
Seq primer: T7  
Class: BAC ends.  
FEATURES  
source Location/Qualifiers  
1..720  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RPI1-429A8"  
/sex="male"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACe3.6 vector at EcoRI sites"  
ORIGIN  
Query Match 9.9%; Score 197.4; DB 10; Length 720;  
Best Local Similarity 75.7%; Pred. No. 8.2e-23;  
Matches 284; Conservative 0; Mismatches 86; Indels 5; Gaps 3;  
QY 1147 GTGTGACAAAGAGAGGCAATTCAGTGTGAATTTGTCAGAGAGTATTGATTATGT 1206  
DB 656 GTGAGAGGGGAGCGGCTGACGCGCAGAGGTTGGTGGCTGACAGAGACATGAGT 597  
QY 1207 TTTCTCAGATCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1266  
DB 596 GTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 537  
QY 1267 TGGAGTACAGTGGCTGTGCTGCTCACTGCAACCTTGGCTCCAGGTTCAAGCAT 1326  
DB 536 TGGAGTGAAGTGG--TGCAATCTTGGCTCACTGCAACCTTGGCTCCCGGTTCAAGCAT 478  
QY 1327 TCTCCG--TCAGCTTCCCGAATAGCTGGATTACAGGCGCATG--ACCAATAGCTTA 1382  
DB 477 TCTCTGCTCAGCTCCCGAGTAGCTGGATTACAGTATGATGCAAAAGCCCGGCTA 418  
QY 1383 ATTTTGTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1442  
DB 417 ATTTTGTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 358  
QY 1443 CTGACTTCAAGTATCCACCCACTCAGCTCCCAAGACATGAGGATTAACAGCATAGGC 1502  
DB 357 CTGACTTCAAGTATCCCGCACTCGGCTCCCAAGTGTGAGATTAACAAAGTAGAC 298  
QY 1503 CACCGTCCCAAGCTT 1517

Db 297 CACCATGCCGCGCT 283

RESULT 34  
LOCUS A0471471/c 660 bp DNA linear GSS 23-APR-1999  
DEFINITION CITBI-B1-2600N15.TF CITBI-B1 Homo sapiens genomic clone 2600N15,  
genomic survey sequence.  
ACCESSION A0471471  
VERSION A0471471.1 GI:4655125  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo  
1 (bases 1 to 660)  
Zhuo, S., Adams, W.D., Niernan, W., Malek, J., Shizuya, H., Simon, M. and  
Venter, J.C.  
TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready  
Map Building  
JOURNAL Unpublished (1997)  
COMMENT Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeet@igf.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M33-21  
Class: BAC ends.

FEATURES  
source location/Qualifiers  
1..660  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="2600N15"  
/sex="male"  
/cell\_type="sperm"  
/clone\_lib="CITBI-B1"  
/note="Vector: pBelBAC1; Site\_1: EcoRI; Site\_2: EcoRI;  
Caltech Human BAC Library D"

ORIGIN  
Query Match 9.9%; Score 197.2; DB 9; Length 660;  
Best Local Similarity 77.6%; Pred. No. 9e-23;  
Matches 264; Conservative 0; Mismatches 73; Indels 3; Gaps 2;

1219 TTTTATTTTATTTTATTTTGAAGAGTCTACTTTGTACACCCAGGCTGGAGTACAGT 1278  
|||||  
583 TTTTCTTTTATTTTATTTTGAAGAGGCTCTCTGTACACCCAGGCTGGAGTACAGT 524  
|||||  
1279 GGTGAGTCTGGGCTACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG--TCA 1336  
|||||  
523 GC-GTATCTCGGCTACTGCAATCTCTGCTCTGATGGGTTCAATGATTTCTTGCTCA 465  
|||||  
1337 GCTTCCGGAATAGCTGGAGTTACAGGCGCATGCAACCAATGCTTAATTTTGTATTTT 1396  
|||||  
464 GTCCTCCGAGTAGCTAGACTACAGAGTGGGCGCATCAAGCTGGGCTAATTTTGTATTTT 405  
|||||  
1397 AGTAGAGACAGAGTTTCCGCAATGTGACCAAGGCTTGCTTGAACCTCTGACTTCAGGTGA 1456  
|||||  
404 GGTAGAGACAGAGGTTTCCACCACTTGGCCAGGCTGGCTTGAACCTCTGACTTCAGGTGA 345  
|||||  
1457 TCACCACTCAAGCTCTCCCAAGCACTGGAGTTTACAGGCGATGAGCAACGTCGACGCC 1516  
|||||  
344 TCACCACTCAAGCTCTCCCAAGGAGTGGAGTTTACAGGCGATGAGCAACCAATGCCGGA 285  
|||||  
1517 TGTTCCTCAGATCTGTATTTTGTTCGAAAGCTTCAAT 1556  
|||||

Db 284 GAAATTTTTCATTTGGCTTAAGACTAAACTGTATTT 245

RESULT 35  
LOCUS AA832175 410 bp mRNA linear EST 18-MAR-1998  
DEFINITION OC91C04.B1 NCI-CGAP GCBI Homo sapiens cDNA clone IMAGE:1357062 3'  
similar to contains Alu repetitive element, mRNA sequence.  
ACCESSION AA832175  
VERSION AA832175.1 GI:2905274  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
1 (bases 1 to 410)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgs@bbs-rtmail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E.S. Consortium/BLND at:  
www.bio.lnl.gov/bbrp/image/image.html  
Insert Length: 1357 Std Error: 0.00  
Seq primer: -40m3 fwd. Err from Amersham  
High quality sequence stop: 402.

FEATURES  
source location/Qualifiers  
1..410  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1357062"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP GCBI"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-GGTACCAACTGTAAGGAGGAGCGCGCTCATTTTATTTT-3'  
]. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN  
Query Match 9.8%; Score 196.8; DB 1; Length 410;  
Best Local Similarity 77.6%; Pred. No. 1.2e-22;  
Matches 264; Conservative 0; Mismatches 72; Indels 4; Gaps 2;

1220 TTTTATTTTATTTTATTTTGAAGAGTCTACTTTGTACACCGAGGCTGGAGTACAGT 1279  
|||||  
1 TTTTATTTTATTTTATTTTGAAGAGGAGTCTCTGTGCTGAGGCTGAGGTGAGTGG 60  
|||||  
1280 CTGTGGTCTGGGCTCACTGCAACTCTGCGCTCCAGGTTCAAGGATTTCTCTG--TCAG 1337  
|||||  
61 CAGAGTCTGGGCTACTGCAACCTCCGCTCCAGGTTCAAGTATTTCTCTGCGCTGAG 120  
|||||  
1338 CTTCGGAATAGCTGGAGTTTACAGGCGCATGCAACCAATGC--CTAATTTTGTATTTT 1395  
|||||



intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transfection of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

## ORIGIN

Query Match 9.8%; Score 196.6; DB 6; Length 634;  
Best Local Similarity 79.8%; Pred. No. 1.1e-22;  
Matches 269; Conservative 0; Mismatches 64; Indels 4; Gaps 3;

QY 1210 CTCAGATCTTTTATTTTATTTTGTGAAACAGATCTCACTTTGTCACCCAGGCTGG 1269  
DB 353 CTGTAACCATTTCTTTTCTTTTGTGATGAGTCTCACTGTGCGCCAGGCTGG 294  
QY 1270 AGTACAGTGGCTGTGCTCTGCTCACTGCAACTTGTCTCCAGTTCAAGCATTTCT 1329  
DB 293 AGTGAAGTAC-ACAACTCGGCTCACTGCAACTCTGCTCTCGGTTCAAGCATTTGT 235  
QY 1330 CCTG--TCAGCTTCCGAAATGCTGGATTAACAGGCGCATGACCA-CCATGCTTAATT 1386  
DB 234 CTGCTCTACGCTCTGAGTAGCTGGATTATGAGCATGTGCCCATGCGCCGCTTAATT 175  
QY 1387 TTGATTTTATGAGAGACAGATTCGCAATGTTGACAGGCTTGCCTTGAATCTCTGA 1446  
DB 174 TTGATTTTATGAGAGAGCGGGTTTCAACATATTGATCAGGTTGCTCGAATCTCTGA 115  
QY 1447 CTTGAGGTATCCACCCACTTCAAGCTCCCAAGCACTGGATTAACAGCATGAGCAC 1506  
DB 114 CCTCAGGTATCCACCCGCTCAAGCTCCCAAGGTCCTGGATTACAGCGTGAAGCAC 55  
QY 1507 GTGCCAGCTGTTTCTCAGATCCGTATTTGTTTC 1543  
DB 54 ATGCTCAGCTTGTGAGCCATTTTCAATGATGATGC 18

RESULT 38  
A0506289 732 bp DNA linear GSS 29-APR-1999  
LOCUS RPCI-11-299L10.TV RPCI-11 Homo sapiens genomic clone  
DEFINITION RPCI-11-299L10, genomic survey sequence.  
ACCESSION A0506289  
VERSION A0506289.1 GI:4711036  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 732)  
Zhaoc,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter,J.C.  
TITLE Use of BAC End Sequences from library RPCI-11 for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other\_GSSs: RPCI-11-299L10.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@igf.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page: [http://www.tigr.org/tcdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: 17  
Class: BAC ends.

## FEATURES

## source

Location/Qualifiers  
1..732

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7614705"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-299L10"  
/sex="Male"  
/cell\_type="lymphocytes"  
/clone\_1ib="RPCI-11"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; RPCI11 Human Male BAC Library"

## ORIGIN

Query Match 9.8%; Score 196.6; DB 9; Length 732;  
Best Local Similarity 75.0%; Pred. No. 1.1e-22;  
Matches 285; Conservative 0; Mismatches 90; Indels 5; Gaps 3;

QY 1201 TTATGTTTTCAGATCTTTTATTTTATTTTGTGAAACAGATCTCACTTTGTAC 1260  
DB 237 TTCCTTAAGTTTGCTTTTATTTTATTTTCTTGAACAGATTTGCTTGAC 296  
QY 1261 CCAGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACTTGTCTCCAGTTCA 1320  
DB 297 CCAGGCTGAGTACAGTGG-TCATCTTGTGCTCACTGCAACTTGTCTCCAGTTCA 355  
QY 1321 AGCGATTCCTCG--TCAGCTTCCGAAATGCTGGATTACAGGCGCATGACCA-- 1376  
DB 356 AGCAATTCCTCGCTTGTGCTTCCGATTAAGCTGGGTTACAGGACATGACCAAC 415  
QY 1377 TCCCTAATTTTGTATTTTATGATAGACAGATTTGCGCATTTGACCAAGCTTGCCT 1436  
DB 416 TGGCTAATTTTGTATTTCTTATGATAGATGGGTTTACCAATTTGGCCAGGCTGCTT 475  
QY 1437 GAATCTCGACTTGAAGTATGATCAACCACTGAGCTCCCAAGACATGGATTACAGG 1496  
DB 476 GAATCTCGACTTGAAGTATGATCAACCACTGAGCTTCCCAAGTCTGATTTACAGG 535  
QY 1497 ATGAGCAACCTGCCCAGGCTGTTTCTCAGATCTGTATTTTGTGAACTTTCA 1556  
DB 536 ATGAGCACTGACCCCGGCAAGTTTGTCAATTTGATTAACCTTTCAAGAAACACTT 595  
QY 1557 TCTATCTTCTTATTTATTTT 1576  
DB 596 TTGGTTTGTGATTTT 615

RESULT 39  
AG182560 669 bp DNA linear GSS 09-JAN-2002  
LOCUS AG182560/c Pan troglodytes DNA, clone: RP43-055M1.1.TV, genomic survey sequence.  
DEFINITION AG182560  
ACCESSION AG182560  
VERSION AG182560.1 GI:16712240  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Pan.  
REFERENCE 1  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE BAC end sequences of library RPCI-43  
JOURNAL Unpublished  
COMMENT 2 (bases 1 to 669)  
REFERENCE  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.





JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764

## COMMENT

Neuberger, Germany  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Sequenced by Agowa (Berlin/Germany) within the cDNA sequencing Consortium of the German Genome Project.  
This clone (DKFZp586G2219) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.  
Please contact RZPD for ordering:  
<http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp586G2219>  
Further information about the clone and the sequencing project is available at <http://mips.gsf.de/projects/cdna/>.

## FEATURES

## source

1. .3110  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="RZPD:DKFZp586G2219"  
/db\_xref="taxon:9606"  
/clone="DKFZp586G2219"  
/issue\_type="uterus"  
/clone\_lib="586 (synonym: hute1). Vector pSport1; host DH10B; sites NotI + SalI/MluI"  
/dev\_stage="adult"  
/note="culteln 1"  
1. .3110  
/gene="DKFZp586G2219"  
51. .1223  
/gene="DKFZp586G2219"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB66851.1"  
/db\_xref="GI:12053329"  
/db\_xref="GOA:O9NMN1"  
/db\_xref="UniProt/Swiss-Pro:O9NMN1"  
/translation="MNGTRNMTLYDVHPEDDAASVDLRUTLOGELTNGDEHIAO KAGRTYAMVSSHSGHSLASLVESHDSHERITVYLKRGSDKQIHAKINQIKSE VOYIGARNCLQKREDISSKLDRLGDSLRHQLQVLEKRGDSQSPALYSPP E VDTCTINDEVSLEKTVODLALQEAQRQHSQCVAFETVLTRYRBAEQSNVALORE EDRVKEAEVEGLQRIILGMETHEHALLAKREGEVAFLEELSRNNADQERAEKAT LEKEVAGREKTIHHH.DDMK.KSQQRVROMI.EIOLNSKAVTOSKXATIOELKIKYLE AENLEMDHMEHLIEKQISKGNFSTQARAKTENPISIRISKPSKPMPIRVET"

## CDS

## gene

## ORIGIN

Query Match 9.8%; Score 196.2; DB 4; Length 3110;  
Best Local Similarity 69.9%; Pred. No. 9e-23;  
Matches 293; Conservative 0; Mismatches 123; Indels 3; Gaps 2;  
1193 GATTGATTATGTTTCTCAGATCTTTTATTTTATTTTGAACAAGTCTCAC 1252  
1245 GAGGTCATCATGTTTCTTTTATTTTATTTTCTGAGACAGATCTCAC 2484  
1253 TTTTTCACCCAGGCTGGAATACAGTGGCTGCTCGCTCAGCAACCTGCTCC 1312  
2485 TCTGTACCCAGGCTGGAATGAGTGC-GCAGTCTCAGCTCACTGCAACTCTGCTCC 2543  
1313 CAGGTTCAAGCAATTCCTG--TCAGCTCCCGAATAGCTGGATTACAGGCGCATGCA 1370  
2544 CGGATTTAAGCATTTCTTCTGCTCAGCTCCGAGTACTGGGATACAGGTGCCAC 2603  
1371 CCACCATGCTAATTTTGTATTTTATGAGACAGAGTTTGCAGTGTGACAGGCT 1430  
2604 CACGCTGCTGATTTTGTATTTTATGAGATGGGGTTTCAACATCTGGCTAGGCT 2663  
1431 TGCCTTGAATCTCTGACTTCAAGTATCCACCACTCAGCCCTCCAAAGCACTGGGATT 1490  
2664 GGTCTGGAATCTCTGACTCAGGATGATCCACCACTCGGCTTCCAAAGTGTAGGATT 2723  
1491 ACAGGATGAGCCACGCTGAGCCAGCTGTTTCTCAGATCTGATTTGTTTGAAGCC 1550  
2724 ATAGGCTTGAAGTACTGTGCGCCGACCAAGTGTCTTTTGAAGGCTCTTCTTACAGCC 2783  
1551 TTTCAATTTATCTTCTTATTTATTTGGAAGTAGTACACCTAAGTAAAGTTTAA 1609  
1551 TTTCAATTTATCTTCTTATTTATTTGGAAGTAGTACACCTAAGTAAAGTTTAA 1609

Db 2784 TTGAAAGTAGATAGGATGAGTACAGTATGTACTATGAGATCAGAAAAATTCAAAACAA 2842

## RESULT 42

## LOCUS

AO885682 510 bp DNA linear GSS 09-NOV-1999

HS\_5525\_A2\_G06\_T7A\_RPCT-11 Human Male BAC Library Homo sapiens

DEFINITION HS\_5525\_A2\_G06\_T7A\_RPCT-11 Human Male BAC Library Homo sapiens

ACCESSION AO885682

VERSION AO885682.1 GI:6317149

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 510)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## COMMENT

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics (<http://info@resgen.com>). BAC end Web Server: <http://www.hbec.washington.edu>  
Plate: 9293 row: M column: 12  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 510.  
Location/Qualifiers

## FEATURES

## source

1. .510  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=9293 Col=12 Row=M"  
/sex="male"  
/clone\_lib="RPCT-11 Human Male BAC Library"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

## ORIGIN

Query Match 9.8%; Score 196; DB 9; Length 510;  
Best Local Similarity 78.0%; Pred. No. 1.5e-22;  
Matches 273; Conservative 0; Mismatches 72; Indels 5; Gaps 3;  
1216 TCTTTTATTTTATTTTATTTTGAACAAGTCTCACTTGTGACCCAGGCTGAGTACA 1275  
104 TCTTTTCTTTTATTTTATTTTGAAGAGAGTCTGCTCTGTGACCCAGGCTGAGTACA 163  
1276 GTGCTGTGCTGCTGCTGCTGCAACCTCTGCTCCAGGTTTCAAGGATCTCTGCTG-- 1333  
164 ATGGC-AGGATCTGAGCTCAGTCACTGCAACTCTGTCTCCAGGCTCAAGCAATTTCTCTGCC 222  
1334 TCAGCTTCCGGAATGCTGGGATTTACAGGCGGATG--ACGACCATGCTTAAATTTTGA 1391  
223 TCAGCTTCCGGAATGCTGGGATTTACAGGCGGATG--ACGACCATGCTTAAATTTTGA 282  
1392 TTTTATGAGACAGAGTGTGCGCATGTTGACCAAGGCTTGCTGAACTCTGCACTTCA 1451

|||||  
Db 283 TTTTATGATGAGAGCGAGGTTTCCACATGTGGCCAGGCTGTCTCAAACTCCGACCTCA 342  
1452 GGTGATCCACCCACCCAGCCTCCCAAGACACTGGATTAACAGGACATGAGCCACCGCTGCC 1511  
Db 343 TGTGATTTGCTGCGCTTGGCTTCCAAAGTTGGAGATTACAGGCAATGAGCCACCATGGCC 402  
1512 CAGCGTGTCTCAGATCCTGATTTGTTCTGAAGCCTTCATTTCTAT 1561  
Db 403 CAGCGTTTTTTTCTTTCTTTCTTTCTTTTGAATCAACTATTGTTT 452

RESULT 43  
AQ379603 724 bp DNA linear GSS 20-MAY-1999  
LOCUS RPCI11-163F3.TU RPCI-11 Homo sapiens genomic clone RPCI-11-163F3,  
DEFINITION genomic survey sequence.  
ACCESSION AQ379603  
VERSION AQ379603.1 GI:4350626  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 724)  
Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and  
Venter,J.C.  
Venter,J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
Other GSSs: RPCI11-163F3.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@igr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@edj.med.bufileo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering>) or from  
Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page:  
[http://www.tigr.org/cdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Classes: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..724  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7562330"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-163F3"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_lib="RPCI-11"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"

ORIGIN  
Query Match 9.8%; Score 196; DB 9; Length 724;  
Best Local Similarity 74.4%; Pred. No. 1.4e-22;  
Matches 287; Conservative 0; Mismatches 95; Indels 4; Gaps 3;  
1206 TTTTCTCAGATCTTTTATTTTATTTTGAACAGAGTCTCACTTGTACCCAGG 1265  
Db 324 TTGTAATATTGTTTGTGTTTGTGTTTGAAGCGAGTCTGCTCTGCGCCAGG 383  
QY 1266 CTGAGATACAGTGGCTGTGCTGCTACTGCAACCTCTGCTCCAGGTTCAAGCA 1325  
Db 384 CTGAGATGACGTGC-ACAATGTGGCTCACTGCAACCTCCAGGTTCAAGTA 442

QY 1326 TTCTCTCG--TCAGCTCCCGAATAGCTGGATTACAGCGCATGACCA-CCATGCCTA 1382  
Db 443 TTCTCTGACTCAGCCTCCCAAGTAGCTGGATTTCAGGCACACACCATGATGCTGA 502  
QY 1383 ATTTTGTATTTTATAGTAGAGACAGATTGGCCATGTTGACCAAGCTTGCTTGAATC 1442  
Db 503 ATTTTGTATTTTATAGTAGAGACAGATTGGCCATGTTGACCAAGCTTGCTTGAATC 562  
QY 1443 CTGACTTCAGGTGATCCACCACCTCAGCTCCCAAGACATGGATTACAGCAAGAGC 1502  
Db 563 CTGACCTCAAGATGATCTGCCACTCTGCTCCAAAGTGAATTACAGGCTTGAGC 622  
QY 1503 CACGTCGCCAGCCTGTTTCTCAGATCCTGATTTGTTCTGAAGCCTTCATTTCTATC 1562  
Db 623 CACCATGTCACAGCTGTTTGTATATTTGTTGAATTTGGCTCTAGTGATTTCTGGC 682  
QY 1563 TTCTTATTCATTTTGAAGTAGTACA 1588  
Db 683 TAAGATTTAAGTGTCTTAGTTGAA 708

RESULT 44  
AQ075668 600 bp DNA linear GSS 20-AUG-1998  
LOCUS CIT-HSP-236713.TF CIT-HSP Homo sapiens genomic clone 236713,  
DEFINITION genomic survey sequence.  
ACCESSION AQ075668  
VERSION AQ075668.1 GI:3437324  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 600)  
Adams,M.D., Rounley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., Shituya,H., Simon,M. and  
Venter,J.C.  
Venter,J.C.  
Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building  
Unpublished (1998)  
Other GSSs: CIT-HSP-236713.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamas@igr.org  
Clones are available from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC  
end search page:  
[http://www.tigr.org/cdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: M13-21  
Classes: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..600  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="236713"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_lib="CIT-HSP"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

ORIGIN  
Query Match 9.8%; Score 195.8; DB 9; Length 600;  
Best Local Similarity 80.0%; Pred. No. 1.6e-22;  
Matches 268; Conservative 0; Mismatches 62; Indels 5; Gaps 3;  
1194 TATTTGATTAATGTTTCTCAGATCTTTTATTTTATTTTGAACAGAGTCTCACT 1253

	Db	242	TCTTTTCTTTTCTTTCTTTCTTTCTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCAGT	301
Oy		1254	TTGTGACCAGCGCTGGAGATACAGTGCCTGTGCTCGCGCTACCTGCAACTCTTCCCTCC	1313
Db		302	CTGTACCCCGGGCTGGAGTGACATCG--TGCGCTCTTGGCTCACAAGCAACTCTGCCCTCG	360
Oy		1314	AGGTTCAAAGCATTTCTCCGT--TGAGCTTCCGAATATGCTGGGATTTCAGGCGCATGCAC	1371
Db		361	GGGTTCAAGTATTTCTCTGCTCCTAGCTCCTCTGAGTAGCTGGAGATTACAGGCATGTACCA	420
Oy		1372	CACCA--TGCCTAATTTTGTATTTTGTATAGTACAGACAGAGTTTGGCATGTTCAGCACGC	1429
Db		421	CCACAGCTCGGTAAATTTTTTTTATTTTATTTAGTAGAGCGGGGTTTACCATATGTGGTCAAGC	480
Oy		1430	TTGCTTGAATCTCTGACTTCAAGTGATCAACCACTCAAGCTTCCAAAAGCATTTGGAT	1489
Db		481	TGGTCTTGAACTCTCGAACCTCAAGGATTCACCAACTTGGCCCTCCCAAAAGTCTGGGAT	540
Oy		1490	TACAGGATGAGGCACCGGTGCCCCAAGCCGTTTTCC	1524
Db		541	TGCAAGCGGTGAGCCACTGACCTGGTCTCTTT	575
RESULT_45				
LOCUS		CR859293	3015 bp	mRNA linear HTC 12-NOV-2004
DEFINITION		Pongo pygmaeus mRNA; cDNA DKFZp468N155 (from clone DKFZp468N155).		
VERSION		CR859293		
KEYWORDS		CR859293.1 GI:55729479		
SOURCE		HTC.		
ORGANISM		Pongo pygmaeus (orangutan)		
		Pongo pygmaeus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarchontoglires; Primates; Catarrhini; Hominoidea; Pongo.		
REFERENCE		1 (bases 1 to 3015)		
AUTHORS		Anseuge,W., Krieger,S., Regiert,T., Ritzmuller,C., Schwager,B., Wewes,H.W., Weil,B., Amdt,C., Osanger,A., Fobo,G., Han,W. and Wiemann,S.		
CONSRMT TITLE		The German CDNA Consortium Direct Submission		
JOURNAL		Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764		
COMMENT		Neuberberg, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp468N155) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: <a href="http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp468N155">http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp468N155</a> Further information about the clone and the sequencing project is available at <a href="http://mips.gsf.de/projects/cdna/">http://mips.gsf.de/projects/cdna/</a> .		
FEATURES		location/Qualifiers		
source		1..3015		
		/organism="Pongo pygmaeus"		
		/mol_type="mRNA"		
		/db_xref="taxon:9600"		
		/clone="DKFZp468N155"		
		/tissue_type="heart"		
		/clone_lib="468 (synonym: phrt1). Vector pSport1_Sfi; host DH10B; sites SfiI + SfiIB"		
		/dev_stage="adult"		
		/note="hypothetical protein (Homo sapiens)"		
		1..3015		
		/gene="DKFZp468N155"		
		95..1996		
		/gene="DKFZp468N155"		
		/codon_start=1		
		/product="hypothetical protein"		
		/protein_id="CAH91471.1"		
		/db_xref="GI:55729480"		
gene				
CDS				

/tiansatlon="MESGEPMJALUJVENNEMOLLVNOALLOILEKISOPVAVAV  
 I/GRKGTAKTMMHJLGAONHGPFLGVSVEETKIMMOLCPHSGKPHNVLITDEBEG  
 DYKEDKPKNDKSIPIALVJLSTETPLYSNHNTHOALBOHVAYTELTIKAKSSRP  
 DAVDSTIEVSFPEDPIVTRPPTLEKANGHRTDEBELNALKILOGNRPVOTSN  
 LPERCIRREFPRKCFORPNDKOLIANIKRVSOKDPKRPQBOTINPFSIITHA  
 RAKTILREGIIVTNGILGTILAVYVENAVGAPCEANAVITLAORNSAVOGAARV  
 SOOMARVAFEDNITLOLMDHACERBAIAPFMESFEDENOEPORKEPMTANNGK  
 DILNEESVYFOCAKNEILSKGAMESISASFSVAGKCIYMERKEERLEDOYVAP  
 RGVYAKVEFORSFOSOVITESSILOSDALINDREKAVAVDRAKKEARELLOK  
 LOBOOOOAMADKSLKENIAJOLKULOWREBOLRPOIMMLLEHTOKVONDMHEGFX  
 KFEENNAEISOFKRMIDITKNDTPITATLIDLADELTAIVTASAPKILGHVKGVS  
 LFFKIKGLHP"

Query Match	9.8%;	Score 195.8;	DB 4;	Length 30.5;
Best Local Similarity	70.0%;	Pred. No. 1.1e-22;		
Matches 292;	Conservative	0;	Mismatches 122;	Indels 3; Gaps 2;

QY	1102	CATGAAAATGCTTCAGAACCCCAATATPAAAGAGATTTTTCAGCATGTGCACAAAAGAGA	1161
Db	1969	CTTTAAAAGCATAGCTCCCTCTTTAAAGATATTATGATTTGACATATATGCTTTGGA	2028
QY	1162	GGCCATTTTCAGATGTGAAATTTGTTCAGAGAATATTTGATATGTTTCTCAATCTTTT	1221
Db	2029	CTATTTTTCGATCTGTATGTTTTCATTTTCATTCAGCAAGTTTTTTTGTGTTGTTT	2088
QY	1222	TATTTTATTTTTTTTTGAAAACAGTCTCACTTTGTCACCAGGCTGGAATACAGTGGCT	1281
Db	2089	TGTTTTGTTTTTTTTATCAGACAGAGCTTACTCTGTGTGCTTGAAGCTGGAATACAGTGG - T	2147
QY	1282	GTGATCTCGGCTACCTGCACACCTCTGCTGCCAGGTCAAGCCATTTCTCTGTCAGCTTC	1341
Db	2148	GCAATCTCAGTCACTGCAACTTCTGCTCCCGGTTTCAAGAGATTTACCTGCTCAGCC	2207
QY	1342	CCGAAATAGCTGGAGTTTACAGGCGCATCCACCACCATGC - - CTATTTTGTATTTTAACT	1399
Db	2208	CTTAGTAGCTGGAGTTTATAGGTGCAACACACACACCAGGTATTTTTGTATTTTAACT	2267
QY	1400	AGAGACAGAGTTTGCACATGTTTACACAGGCTTGCTTTGAATCTCTGACTTCAGGTATCC	1459
Db	2268	AGAGATGGGGTTTACACATGTTTGCCACAGGCTGCTCTGAACTTTTTCACCTCAATATGATCC	2327
QY	1460	ACCCACCTCBAACCTCCCAAAAGCACTGGGATTTACAGGAGATGAGCAGCGTGCCAGCC	1516
Db	2328	ACCGGCTCAGACCTCCCAAAATGCTGGGATTTACAGGAGATGAGCAGCAGCGCCAGCC	2384

Search completed: January 22, 2006, 21:09:07  
Job time : 5428.54 secs

gene	1. .3015
	/gene="DKFZp468N15"
CDS	95. .1996

```

/cocon_start=1
/product="hypothetical protein"
/protein_id="CAH91471.1"
/db_xref="GI:55729480"

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 : Search time 6808.48 Seconds  
(without alignments)  
16706.187 Million cell updates/sec

Title: US-09-728-552A-3\_COPY\_19000\_21000

Perfect score: 2001

Sequence: 1 aatgcttgattccaagtc.....caggacatcggtgcatagc 2001

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database :

1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_hg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2001	100.0	40917	6	AX033911 Sequence
2	2001	100.0	80622	8	AF222855 Homo sapi
3	1974.2	98.7	41008	6	AX033912 Sequence
4	1974.2	98.7	69058	8	AF222856 Homo sapi
5	1974.2	98.7	80155	8	AF042484 Homo sapi
6	1974.2	98.7	80202	8	AF222854 Homo sapi
7	1974.2	98.7	112442	8	AL355340 Homo sapi
8	1974.2	98.7	176432	14	AL391648 Homo sapi
9	670.4	33.5	71032	8	AC084084 Homo sapi
10	218	10.9	27423	8	AP006290 Homo sapi
11	218	10.9	13315	14	AC136997 Homo sapi
12	218	10.9	161953	8	AC130304 Homo sapi
13	216.6	10.8	98340	8	AL353716 Homo sapi
14	215.8	10.8	58330	14	AL353694_3 Homo sapi
15	215.2	10.8	165007	8	AC019047 Homo sapi
16	215	10.7	172296	14	AP002889 Homo sapi
17	214	10.7	89323	8	AP000221 Homo sapi
18	214	10.7	100000	8	AP000085 Homo sapi

19	214	10.7	100000	8	AP000137 Homo sapi
20	214	10.7	340000	8	AP001693 Homo sapi
21	213.8	10.7	164857	8	AC110054 Homo sapi
22	213.6	10.7	72048	6	AR659559 Sequence
23	213.6	10.7	72048	6	AR659643 Sequence
24	213.6	10.7	168751	14	AC068969 Homo sapi
25	213.6	10.7	170883	8	AC067805 Homo sapi
26	213.4	10.7	175752	14	AC091721 Homo sapi
27	213.4	10.7	175808	8	AC146040 Pan trogl
28	213.2	10.7	42673	8	AC138517 Homo sapi
29	212.8	10.6	179836	8	AL3590822 Homo sapi
30	212.8	10.6	213432	14	AC068198 Homo sapi
31	212	10.6	130080	8	AP003160 Homo sapi
32	212	10.6	163848	8	HS105613 Homo sapi
33	211.8	10.6	70446	8	AC107299 Homo sapi
34	211.8	10.6	118135	8	AC117477 Homo sapi
35	211.8	10.6	181505	14	AC025415 Homo sapi
36	211.6	10.6	190264	14	AC161624 Homo sapi
37	211.6	10.6	226170	8	AC008761 Homo sapi
38	211	10.5	44553	8	D88268 Homo sapien
39	211	10.5	178000	8	AP002893 Homo sapi
40	210.8	10.5	87071	8	CR589678 Homo sapi
41	210.8	10.5	107537	8	CR954268 Homo sapi
42	210.8	10.5	138165	8	CR936851 Homo sapi
43	210.8	10.5	164033	8	CR938745 Homo sapi
44	210.8	10.5	175639	8	CR936885 Homo sapi
45	210.8	10.5	207664	8	AC157497 Homo sapi
46	209.8	10.5	159512	14	AC144369 Homo sapi
47	209.8	10.5	166336	8	AL445467 Homo sapi
48	209.8	10.5	175339	8	AC005772 Homo sapi
49	209.6	10.5	61505	8	AL512330 Homo sapi
50	209.6	10.5	162774	14	AC034138 Homo sapi
51	209.6	10.5	173149	14	AL391218 Homo sapi
52	209.4	10.5	70248	14	AC103795 Homo sapi
53	209.4	10.5	149466	8	HS42784 Homo sapi
54	209.4	10.5	182329	8	AP001646 Homo sapi
55	209.2	10.5	145705	8	AC090255 Homo sapi
56	209.2	10.5	158458	8	AC007395 Homo sapi
57	209.2	10.5	170394	14	AC024371 Homo sapi
58	209	10.4	100731	8	AL133347 Homo sapi
59	209	10.4	107731	8	AC005476 Homo sapi
60	208.8	10.4	137143	8	AL451136 Homo sapi
61	208.8	10.4	160342	14	AC025492 Homo sapi
62	208.8	10.4	165434	14	AC087505 Homo sapi
63	208.8	10.4	165446	14	AC069217 Homo sapi
64	208.8	10.4	190712	14	AC165405 Homo sapi
65	208.6	10.4	106008	8	HS506 Homo sapi
66	208.6	10.4	155729	8	CNS01DMN Homo sapi
67	208.6	10.4	178369	14	CT005241 Homo sapi
68	208.6	10.4	192097	8	AC008743 Homo sapi
69	208.6	10.4	196015	14	AC150024 Homo sapi
70	208.4	10.4	6280	8	HS805888 Homo sapi
71	208.4	10.4	137288	8	AL161909 Homo sapi
72	208.4	10.4	148151	8	AL513533 Homo sapi
73	208.4	10.4	154323	14	AC024603 Homo sapi
74	208.4	10.4	159367	14	AC044812 Homo sapi
75	208.4	10.4	164883	8	AC090282 Homo sapi
76	208.4	10.4	183113	8	HS24387 Homo sapi
77	208.4	10.4	187990	14	AL590805 Homo sapi
78	208.2	10.4	46919	8	AF165141 Homo sapi
79	208.2	10.4	73463	8	AL161785 Homo sapi
80	208.2	10.4	121628	8	AC117517 Homo sapi
81	208.2	10.4	171951	8	CNS0502DA Homo sapi
82	208.2	10.4	192324	8	AC096920 Homo sapi
83	208.2	10.4	213447	14	AC083833 Homo sapi
84	208	10.4	87034	8	AC004587 Homo sapi
85	208	10.4	88955	14	AY442342 Homo sapi
86	208	10.4	143796	14	AC006411 Homo sapi
87	208	10.4	170245	8	HS109714 Homo sapi
88	208	10.4	173701	8	AC018946 Homo sapi
89	208	10.4	223972	14	AC027478 Homo sapi
90	207.6	10.4	71179	8	AL136110 Homo sapi
91	207.6	10.4	108906	8	AC022088 Homo sapi

92	207.6	10.4	122564	8	HSJ899816	AL050338 Human DNA
93	207.6	10.4	145588	8	AC010000	AC010000 Homo sapi
94	207.6	10.4	162139	8	AC005682	AC005682 Homo sapi
95	207.6	10.4	164018	14	AC074242	AC074242 Homo sapi
96	207.6	10.4	192441	14	AC073404	AC073404 Homo sapi
97	207.6	10.4	208066	8	AC146483	AC146483 Pan trogl
98	207.4	10.4	49616	8	AL365267	AL365267 Human DNA
99	207.4	10.4	159266	14	AC026097	AC026097 Homo sapi
100	207.2	10.4	106377	8	AC004132	AC004132 Homo sapi

## ALIGNMENTS

RESULT 1  
AK033911  
LOCUS AX033911 40917 bp DNA linear PAT 21-SEP-2000  
DEFINITION Sequence 3 from Patent WO9851790.  
ACCESSION AX033911  
VERSION AX033911.1 GI:10280479  
KEYWORDS  
SOURCE .  
ORGANISM unidentified  
unidentified  
unclassified sequences.

REFERENCE  
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.  
TITLE A novel nucleic acid molecule  
JOURNAL Patent: WO 9851790-A 3 19-NOV-1998;  
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART  
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)

FEATURES  
source  
1..40917  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 100.0%; Score 2001; DB 6; Length 40917;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AATGTTGGATTTC	CAAGGTAACAGAGCTGGAAAAACAATGACCTGCGGAGTCTGAGAG	60
DB	19000	AATGTTGGATTTC	CAAGGTAACAGAGCTGGAAAAACAATGACCTGCGGAGTCTGAGAG	19059
QY	61	TTTCAGCGGAGACT	GTGTGAGCTTGTGTTTTCCTGACGTGCAAGTAAATGAGCCGACG	120
DB	19060	TTTCAGCGGAGACT	GTGTGAGCTTGTGTTTTCCTGACGTGCAAGTAAATGAGCCGACG	19119
QY	121	TTTCAGTGAAGCTT	GTGTTCCCTCCCTCAAGGTTACCCACAATTCCTCACTCTCAAG	180
DB	19120	TTTCAGTGAAGCTT	GTGTTCCCTCCCTCAAGGTTACCCACAATTCCTCACTCTCAAG	19179
QY	181	AAAGCCAAAAAAT	TGAATTTAGGGTTTAGGATTTGGTTCTTTATCTATTAACAGATTG	240
DB	19180	AAAGCCAAAAAAT	TGAATTTAGGGTTTAGGATTTGGTTCTTTATCTATTAACAGATTG	19239
QY	241	ATTAATATGTTCT	CCACCGATGTTCTGCTTGTAAACAATACTACTTCTGACACTACAG	300
DB	19240	ATTAATATGTTCT	CCACCGATGTTCTGCTTGTAAACAATACTACTTCTGACACTACAG	19299
QY	301	CATATGACAGAGT	GTACTACCAAGGTAACACAGAAATGGCTGGCCCAATTCCAAAATCC	360
DB	19300	CATATGACAGAGT	GTACTACCAAGGTAACACAGAAATGGCTGGCCCAATTCCAAAATCC	19359
QY	361	TGNAATGATGAGA	GAATTCAGAAATTAATAAGGGATTTCAACAGAGCTGGCTACGAGAT	420
DB	19360	TGNAATGATGAGA	GAATTCAGAAATTAATAAGGGATTTCAACAGAGCTGGCTACGAGAT	19419
QY	421	TGCGACTGTGTCA	TATCTTGTCAATCATACGAGGTGCTGCTGTCTAGCAACTGTCTCA	480
DB	19420	TGCGACTGTGTCA	TATCTTGTCAATCATACGAGGTGCTGCTGTCTAGCAACTGTCTCA	19479

QY	481	CTGCTTCAATTC	CGCTTGCTGCTCTTTAATATCTGCTTTCTCAGCTCAATGGCTTTCTT	540
DB	19480	CTGCTTCAATTC	CGCTTGCTGCTCTTTAATATCTGCTTTCTCAGCTCAATGGCTTTCTT	19539
QY	541	CCCTCTGCACTGC	ATGCTTTCTTGGGTCAACAGCAATGATTTCTTTAGAACTACCTGAT	600
DB	19540	CCCTCTGCACTGC	ATGCTTTCTTGGGTCAACAGCAATGATTTCTTTAGAACTACCTGAT	19599
QY	601	ACTCAAGAGACTA	CAAGACATTTGGGCATTCGACTTCCACTCTCTGGAAAAACAATTTTA	660
DB	19600	ACTCAAGAGACTA	CAAGACATTTGGGCATTCGACTTCCACTCTCTGGAAAAACAATTTTA	19659
QY	661	TGGAAGCCAGGTT	GCCATTAAGTGGCTCTTGAGGTTGTTTCTCAGCCAGGCCCAAGCTT	720
DB	19660	TGGAAGCCAGGTT	GCCATTAAGTGGCTCTTGAGGTTGTTTCTCAGCCAGGCCCAAGCTT	19719
QY	721	TGTGCTTCAAA	CATGAATTTAGAGGCTTCGAACCAATCCACTTTTCAATGGGCTCA	780
DB	19720	TGTGCTTCAAA	CATGAATTTAGAGGCTTCGAACCAATCCACTTTTCAATGGGCTCA	19779
QY	781	CCCAACTGATAAA	AGAACCAATTTGCCATATCTCAATGACCACTTTTCAAGTGGAGTGG	840
DB	19780	CCCAACTGATAAA	AGAACCAATTTGCCATATCTCAATGACCACTTTTCAAGTGGAGTGG	19839
QY	841	TGATGCTGGAAT	GTGGTCAAGCATTTGCCCAACCACTTTTGGCAAAAAAGCTGGAAGCT	900
DB	19840	TGATGCTGGAAT	GTGGTCAAGCATTTGCCCAACCACTTTTGGCAAAAAAGCTGGAAGCT	19899
QY	901	CTGACTGGGGACC	CTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGAAATGGAACCC	960
DB	19900	CTGACTGGGGACC	CTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGAAATGGAACCC	19959
QY	961	GTGTATGATAT	ATATGCTAAAGGGTTGGCTTTATGTTTCTATTCCTTCAAAACCTGTAG	1020
DB	19960	GTGTATGATAT	ATATGCTAAAGGGTTGGCTTTATGTTTCTATTCCTTCAAAACCTGTAG	20019
QY	1021	AATGATATATG	CTGTTTCCCTTTAAAAAATGCAACAATTTGATATGATGCTGTGAT	1080
DB	20020	AATGATATATG	CTGTTTCCCTTTAAAAAATGCAACAATTTGATATGATGCTGTGAT	20079
QY	1081	AGTAATCTCA	CAGATCATGCTCCATGAATAATGTTCAAGAACCAATTAAGAGATTTT	1140
DB	20080	AGTAATCTCA	CAGATCATGCTCCATGAATAATGTTCAAGAACCAATTAAGAGATTTT	20139
QY	1141	AGCCATGTGTGA	CAAAAGAGGCGCATTTCACTGTGAAATTTGTCAGAGAAATATTTGA	1200
DB	20140	AGCCATGTGTGA	CAAAAGAGGCGCATTTCACTGTGAAATTTGTCAGAGAAATATTTGA	20199
QY	1201	TTATGTTTCT	CAGATCTTTTATTTTATTTTGTGAAACAGAGTCTCACTTTGTCA	1260
DB	20200	TTATGTTTCT	CAGATCTTTTATTTTATTTTGTGAAACAGAGTCTCACTTTGTCA	20259
QY	1261	CGAGCTGAGTA	CAAGTGCCTGCTCGGCTCACTGCAACTGCTGCTCCAGGTTTCA	1320
DB	20260	CGAGCTGAGTA	CAAGTGCCTGCTCGGCTCACTGCAACTGCTGCTCCAGGTTTCA	20319
QY	1321	AGCGATTTCT	CTGCTGAGCTTCCCGAAATGCTGGATTAACAGGCGCATGACCACTG	1380
DB	20320	AGCGATTTCT	CTGCTGAGCTTCCCGAAATGCTGGATTAACAGGCGCATGACCACTG	20379
QY	1381	TATATTTTGT	ATTTTATAGAGACAGAGTTTCCGCAATTTGACAGGCTTGCTTGAAC	1440
DB	20380	TATATTTTGT	ATTTTATAGAGACAGAGTTTCCGCAATTTGACAGGCTTGCTTGAAC	20439
QY	1441	TCTGTACTTCA	GCTGATCCACCACCTGAGCTTCCCAAGCACTGGGATTTACAGGATGA	1500
DB	20440	TCTGTACTTCA	GCTGATCCACCACCTGAGCTTCCCAAGCACTGGGATTTACAGGATGA	20499
QY	1501	GGCACCGTGC	CAAGCTGTTTCTCAGATCTGTAATTTTCTGTAAGGCTTCAATTTCTA	1560
DB	20500	GGCACCGTGC	CAAGCTGTTTCTCAGATCTGTAATTTTCTGTAAGGCTTCAATTTCTA	20559
QY	1561	TCTTCTTATTA	CTTTTGAAGATGTAACACTTAAGTAAAGGTTTTTAACAATCAAAATCTT	1620

Db 20560 TCTTCTTATTCATTTTGGAAAGTAGTACACCTAAGTAAGTTTAAACAATCAATATATCTT 20619  
Qy 1621 TGGAAATTCCTCGTTCCTTTCTTATTCCTTCAAAAAAATATGTACATTAAGTATGTT 1680  
Db 20620 TGGAAATTCCTCGTTCCTTTCTTATTCCTTCAAAAAAATATGTACATTAAGTATGTT 20679  
Qy 1681 ATGTTCTTTCAAAATTAATCATTTCTATCTCAGAAATTTATCTCAGGCTTAATGTTAT 1740  
Db 20680 ATGTTCTTTCAAAATTAATCATTTCTATCTCAGAAATTTATCTCAGGCTTAATGTTAT 20739  
Qy 1741 TGAATAGTCTTCACTTCTTGTGATCCAGTTTCTGCTCTTATTTCACTTAAGTCTAAG 1800  
Db 20740 TGAATAGTCTTCACTTCTTGTGATCCAGTTTCTGCTCTTATTTCACTTAAGTCTAAG 20799  
Qy 1801 TGGCTATTAGATAAAGAGCTTGTACAGATTTCTTCCAAATATGCTTATCTTTTAC 1860  
Db 20800 TGGCTATTAGATAAAGAGCTTGTACAGATTTCTTCCAAATATGCTTATCTTTTAC 20859  
Qy 1861 TGCATGCCAGTACAAACCTTTACGTTTGTATTTGATTTCTTATTAACATTCACAGAAATGCG 1920  
Db 20860 TGCATGCCAGTACAAACCTTTACGTTTGTATTTGATTTCTTATTAACATTCACAGAAATGCG 20919  
Qy 1921 TGACTCTCTCTTCTTCCGTAAGGCAATGCCCAAGCAGACGATTTGTAGATATGTACGCA 1980  
Db 20920 TGACTCTCTCTTCTTCCGTAAGGCAATGCCCAAGCAGACGATTTGTAGATATGTACGCA 20979  
Qy 1981 ACAGGCAATGGGTGTCATAGC 2001  
Db 20980 ACAGGCAATGGGTGTCATAGC 21000

RESULT 2  
AF222855 80622 bp DNA linear PRI 17-JUL-2000

LOCUS Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.  
DEFINITION AF222855  
ACCESSION AF222855.1 GI:9246845  
VERSION  
KEYWORDS

SOURCE  
ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 80622)  
AUTHORS Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,  
Irvine,D.V., Saffery,R. and Choo,K.H.

TITLE The 10q25 neocentromere and its inactive progenitor have identical  
primary nucleotide sequence: further evidence for epigenetic  
modification

JOURNAL Genome Res. 10 (6), 832-838 (2000)

PUBMED 10854414  
2 (bases 1 to 80622)

REFERENCE Barry,A.E.

AUTHORS Direct Submission

JOURNAL Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch  
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,  
Parkville, Melbourne, Victoria 3052, Australia

REMARK Genomic sequence from human 10q25.2, clone11b=HC  
FEATURES  
Location/Qualifiers

1. 80622

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="10"

/map="10q25.2"

/clone="HC"

/complement(3..176)

/rpt\_family="Alu"

repeat\_region complement(979..3564)

repeat\_region /rpt\_family="Tigser1"

repeat\_region 2967..3094

repeat\_region /rpt\_family="Alu"

repeat\_region 3664..3934  
/rpt\_family="Alu"  
complement(4920..5181)  
repeat\_region /rpt\_family="Alu"  
complement(6287..6579)  
repeat\_region /rpt\_family="Alu"  
6625..6966  
/rpt\_family="MLT2B2"  
complement(9215..9490)  
repeat\_region /rpt\_family="Alu"  
complement(11280..11565)  
repeat\_region /rpt\_family="Alu"  
complement(12428..12675)  
repeat\_region /rpt\_family="Alu"  
complement(14304..14583)  
repeat\_region /rpt\_family="Alu"  
17503..17794  
/rpt\_family="Alu"  
complement(20250..20533)  
repeat\_region /rpt\_family="Alu"  
complement(23031..23323)  
repeat\_region /rpt\_family="Alu"  
complement(24301..24479)  
repeat\_region /rpt\_family="MIR"  
24951..25250  
/rpt\_family="Alu"  
complement(25987..26342)  
repeat\_region /rpt\_family="THE1"  
27097..27335  
/rpt\_family="MER2"  
27538..27814  
/rpt\_family="Alu"  
27891..28161  
/rpt\_family="Alu"  
29063..29334  
/rpt\_family="Alu"  
complement(31025..31093)  
repeat\_region /rpt\_family="MLT1"  
complement(31817..32352)  
repeat\_region /rpt\_family="MER1"  
complement(32489..32607)  
repeat\_region /rpt\_family="MLT1"  
33039..33316  
/rpt\_family="Alu"  
33469..33760  
/rpt\_family="Alu"  
complement(34527..34590)  
repeat\_region /rpt\_family="MIR"  
34689..34995  
/rpt\_family="Alu"  
complement(36349..36437)  
repeat\_region /rpt\_family="MIR"  
complement(38796..39079)  
repeat\_region /rpt\_family="Alu"  
39441..39693  
/rpt\_family="Alu"  
complement(43176..43237)  
repeat\_region /rpt\_family="MIR"  
complement(445203..45478)  
repeat\_region /rpt\_family="Alu"  
45892..46025  
/rpt\_family="MIR"  
complement(46618..46851)  
repeat\_region /rpt\_family="Alu"  
48623..48939  
/rpt\_family="Alu"  
complement(49048..50698)  
repeat\_region /rpt\_family="L1"  
complement(50710..50972)  
repeat\_region /rpt\_family="Alu"  
complement(51031..51192)  
repeat\_region /rpt\_family="L1"  
complement(51450..51739)

```
repeat_region      /rpt_family="Alu"
                    complement(53270)
                    /rpt_family="Alu"
repeat_region      54261..54808
                    /rpt_family="LTR12"
repeat_region      59460..60054
                    /rpt_family="LTR12"
repeat_region      60349..60633
                    /rpt_family="Alu"
                    complement(60669)
repeat_region      60801
                    /rpt_family="MIR"
                    complement(61684)
repeat_region      63466..63753
                    /rpt_family="MERS"
repeat_region      64063..64328
                    /rpt_family="Alu"
repeat_region      64446..64521
                    /rpt_family="MERS"
repeat_region      65794..65883
                    /rpt_family="MERS"
repeat_region      66772..66951
                    /rpt_family="MER20"
repeat_region      complement(67895)
                    /rpt_family="MER41"
repeat_region      complement(73192)
                    /rpt_family="Alu"
repeat_region      74086..74367
                    /rpt_family="Alu"
repeat_region      75916..76183
                    /rpt_family="Alu"
repeat_region      complement(77771)
                    /rpt_family="Alu"
repeat_region      78059
```

## ORIGIN

Query Match 100.0%; Score 2001; DB 8; Length 80622;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AATGTTGGATTTCACAGGTAACAGAACTGGAAACAACTACGCGGAGTGTGAGG 60
DB 19027 AATGTTGGATTTCACAGGTAACAGAACTGGAAACAACTACGCGGAGTGTGAGG 19086
QY 61 TTTCAGCGGAGACTGGTGACGCTTGTGTTTTCACAGCTGACAGCTGAATGAGCCGAC 120
DB 19087 TTTCAGCGGAGACTGGTGACGCTTGTGTTTTCACAGCTGACAGCTGAATGAGCCGAC 19146
QY 121 TTTCAGTGAAGCTTGTCTCTCTCTCTCAAGGTTACCAATTCAGTTCTCTGAG 180
DB 19147 TTTCAGTGAAGCTTGTCTCTCTCTCTCAAGGTTACCAATTCAGTTCTCTGAG 19206
QY 181 AAGGCCAAAATGAATTTAGAGGTTAGAGTTGAGTTCTTTATCTTTACAGGATG 240
DB 19207 AAGGCCAAAATGAATTTAGAGGTTAGAGTTGAGTTCTTTATCTTTACAGGATG 19266
QY 241 ATATATATGTTCTCCACAGATGTTCTGCTGTGAACAATCACTTCTGACACTG 300
DB 19267 ATATATATGTTCTCCACAGATGTTCTGCTGTGAACAATCACTTCTGACACTG 19326
QY 301 CATATGACAGAGTGTACTACCAAGGTAAACACAGATTGGCTGCCAATTCGAATCCC 360
DB 19327 CATATGACAGAGTGTACTACCAAGGTAAACACAGATTGGCTGCCAATTCGAATCCC 19386
QY 361 TGAACGTAGTGAAGAAATACAGAAATTAATAATAGGGATTCACAGAGCTGGCTACGAG 420
DB 19387 TGAACGTAGTGAAGAAATACAGAAATTAATAATAGGGATTCACAGAGCTGGCTACGAG 19446
QY 421 TGCAGGTGTCAGATCTTGTGCTCATCATACGAGGTCGTGCTCTAGCACTGCTCA 480
DB 19447 TGCAGGTGTCAGATCTTGTGCTCATCATACGAGGTCGTGCTCTAGCACTGCTCA 19506
QY 481 CTGCTTCATTTCTGCTGCTGCTTTAAATACGCTTTTCTCAGCTCAATGGCTTTCTT 540
DB 19507 CTGCTTCATTTCTGCTGCTGCTTTAAATACGCTTTTCTCAGCTCAATGGCTTTCTT 19566
```

```
QY 541 CCTCTGGCAGTCACTTTCTTTGGGTCAACACAGCAATGATCTTTAGAAATCACTGGT 600
DB 19567 CCTCTGGCAGTCACTTTCTTTGGGTCAACACAGCAATGATCTTTAGAAATCACTGGT 19626
QY 601 ACTCAAGGAGCTACAGACATTTGGGATCCACTTCCACTCTCTTTGGAAAAACAATTTTA 660
DB 19627 ACTCAAGGAGCTACAGACATTTGGGATCCACTTCCACTCTCTTTGGAAAAACAATTTTA 19686
QY 661 TGGAAAGCCAGGTTGCCATATGCTCTTGAAGTTGTTGCTCAGCCCAAGGCCCAAGCTT 720
DB 19687 TGGAAAGCCAGGTTGCCATATGCTCTTGAAGTTGTTGCTCAGCCCAAGGCCCAAGCTT 19746
QY 721 TGTGCTTCACACATGAATTAAGAGCTTCAGAAACAAGATCAATTTTCAATGAGCTCTCA 780
DB 19747 TGTGCTTCACACATGAATTAAGAGCTTCAGAAACAAGATCAATTTTCAATGAGCTCTCA 19806
QY 781 CCCAATCTGATTAAGAAACAATTTGCCATATCTCAATGACCACTTTTCAAGTGGGATG 840
DB 19807 CCCAATCTGATTAAGAAACAATTTGCCATATCTCAATGACCACTTTTCAAGTGGGATG 19866
QY 841 TAGATCTGAATGGGTACAGACATTTGCCCAACCAACTTTGCAAAAAAGGCTGGAACT 900
DB 19867 TAGATCTGAATGGGTACAGACATTTGCCCAACCAACTTTGCAAAAAAGGCTGGAACT 19926
QY 901 CTGACTGGGAGCCCTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGATATGAAACCC 960
DB 19927 CTGACTGGGAGCCCTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGATATGAAACCC 19986
QY 961 GTGATGATATATGCTTAAGAGGTTGGCTTTATATGTTTCTATTTCTTCAAAACCTGGATG 1020
DB 19987 GTGATGATATATGCTTAAGAGGTTGGCTTTATATGTTTCTATTTCTTCAAAACCTGGATG 20046
QY 1021 AATAGATATGCTGTTTCCCTTTAAAAATGCAACAATGCTATTAATGATGCTGTAT 1080
DB 20047 AATAGATATGCTGTTTCCCTTTAAAAATGCAACAATGCTATTAATGATGCTGTAT 20106
QY 1081 AGTAATCAAGATCATGCTCCATGAATAATGCTTCAGAACCCATATATAGAGATTTTTT 1140
DB 20107 AGTAATCAAGATCATGCTCCATGAATAATGCTTCAGAACCCATATATAGAGATTTTTT 20166
QY 1141 AGCCATGTGTGCAAAAAGAGGCCATTTCACTGTGTAATTTGTTCAAGAAATATTTGA 1200
DB 20167 AGCCATGTGTGCAAAAAGAGGCCATTTCACTGTGTAATTTGTTCAAGAAATATTTGA 20226
QY 1201 TTATGTTTTCAGATCTTTTATTTTATTTTATTTTATTTTGAACAGAGCTCACTTTGTAC 1260
DB 20227 TTATGTTTTCAGATCTTTTATTTTATTTTATTTTATTTTGAACAGAGCTCACTTTGTAC 20286
QY 1261 CCAGCTGAGATACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTCA 1320
DB 20287 CCAGCTGAGATACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTCA 20346
QY 1321 AGCGATTCCTGTGAGCTTCCGAATAGCTGGATTAACAGGCGCATGACACATGCT 1380
DB 20347 AGCGATTCCTGTGAGCTTCCGAATAGCTGGATTAACAGGCGCATGACACATGCT 20406
QY 1381 TAAATTTTGTATTTTATAGAGACAGAGTTTGGCATGTTGACAGAGCTTGCTTGAAC 1440
DB 20407 TAAATTTTGTATTTTATAGAGACAGAGTTTGGCATGTTGACAGAGCTTGCTTGAAC 20466
QY 1441 TCCGATCTCAGGTATCCACCACTCAGCTCCCAAGCACTGGGATTAACAGCATGA 1500
DB 20467 TCCGATCTCAGGTATCCACCACTCAGCTCCCAAGCACTGGGATTAACAGCATGA 20526
QY 1501 GCCACGTCGCCAGCTGTTTCTCAGATCTGTATTTGTTCTGAAGCTTCAATTTCTA 1560
DB 20527 GCCACGTCGCCAGCTGTTTCTCAGATCTGTATTTGTTCTGAAGCTTCAATTTCTA 20586
QY 1561 TCTTTCTATATCTTTTGAAGATATACCTTAAGTAAGTTTATTAATCAATATCTT 1620
DB 20587 TCTTTCTATATCTTTTGAAGATATACCTTAAGTAAGTTTATTAATCAATATCTT 20646
```



QY 1621 TGGAAATATCCGCTGCTCTTCTATTCCTACAAATATGTCAGTATAGTATGTT 1680  
|  
|  
|  
Db 20647 TGGAAATATCCGCTGCTCTTCTATTCCTACAAATATGTCAGTATAGTATGTT 20706  
|  
|  
|  
QY 1681 ATGTTCTTCAATATATTCATTTCTATCTCAGATTTATCTCATATGCTTAT 1740  
|  
|  
|  
Db 20707 ATGTTCTTCAATATATTCATTTCTATCTCAGATTTATCTCATATGCTTAT 20766  
|  
|  
|  
QY 1741 TGAATAGTCTTCACTTCTGTCATCCAGTTTCTGCTCTTATTTTCACTTAAGTCTAAG 1800  
|  
|  
|  
Db 20767 TGAATAGTCTTCACTTCTGTCATCCAGTTTCTGCTCTTATTTTCACTTAAGTCTAAG 20826  
|  
|  
|  
QY 1801 TGGCTATTGAATTAAGAGCTTGTAACAGATTCCTTCTCCAAATAGTCTTATCTTTTGC 1860  
|  
|  
|  
Db 20827 TGGCTATTGAATTAAGAGCTTGTAACAGATTCCTTCTCCAAATAGTCTTATCTTTTGC 20886  
|  
|  
|  
QY 1861 TGCATGCCAGTACAACTGTTTAACTGTTTGAATTTCTTCAATCAATTCACAGAACTATGC 1920  
|  
|  
|  
Db 20887 TGCATGCCAGTACAACTGTTTAACTGTTTGAATTTCTTCAATCAATTCACAGAACTATGC 20946  
|  
|  
|  
QY 1921 TGAATCTCTCTCTTCTGTAAGCAATGCCCAAGACAGCAATGTTAGATAGTATGATGCA 1980  
|  
|  
|  
Db 20947 TGAATCTCTCTCTTCTGTAAGCAATGCCCAAGACAGCAATGTTAGATAGTATGATGCA 21006  
|  
|  
|  
QY 1981 ACAGGACATGGGTGATAGC 2001  
|  
|  
|  
Db 21007 ACAGGACATGGGTGATAGC 21027  
|  
|  
|

## RESULT 3

AX033912

LOCUS AX033912 41008 bp DNA linear PAT 21-SEP-2000

DEFINITION Sequence 4 from Patent WO9851790.

ACCESSION AX033912

VERSION AX033912.1 GI:10280480

KEYWORDS

SOURCE unidentifed

ORGANISM unclassified

REFERENCE unclassified sequences.

1

Cancilla,M.R., Choo,K.H. and Du,S.D.

A novel nucleic acid molecule

Patent: WO 9851790-A 4 19-NOV-1998;

CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART

DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)

Location/Qualifiers

FEATURES

source

1..41008

/organism="unidentifed"

/mol\_type="unassigned DNA"

/db\_xref="taxon:32644"

## ORIGIN

Query Match 98.7%; Score 1974.2; DB 6; Length 41008;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 AATGTTTGATTTCAAGGTAAACAAGAACTGAAACAACTACTGCGCGAGTCTGAGG 60  
|  
|  
|  
Db 18736 AATGTTTGATTTCAAGGTAAACAAGAACTGAAACAACTACTGCGCGAGTCTGAGG 18795  
|  
|  
|  
QY 61 TTTTCAGCGGAGACTGCTGACGCTGTTGTTTTTCCATGACAGCTGAAATAGCCACG 120  
|  
|  
|  
Db 18796 TTTTCAGCGGAGACTGCTGACGCTGTTGTTTTTCCATGACAGCTGAAATAGCCACG 18855  
|  
|  
|  
QY 121 TTCAGTGAAGCTGTTTCTTCCTCCCTCCTCAAGTTTACCAAACTCTGAGTCTGACG 180  
|  
|  
|  
Db 18856 TTCAGTGAAGCTGTTTCTTCCTCCCTCCTCAAGTTTACCAAACTCTGAGTCTGACG 18915  
|  
|  
|  
QY 181 AAAGCCAAAATGAATTTAGAGGTTTAGAATTTGTTCTTTATCTATTAACAGATTTG 240  
|  
|  
|  
Db 18916 AAAGCCAAAATGAATTTAGAGGTTTAGAATTTGTTCTTTATCTATTAACAGATTTG 18975  
|  
|  
|

QY 241 ATAATATGTTCTCCACAGATGTTCTGCTTGTAACAAATACTACTCTCCGACACTAG 300  
|  
|  
|  
Db 18976 ATAATATGTTCTCCACAGATGTTCTGCTTGTAACAAATACTACTCTCCGACACTAG 19035  
|  
|  
|  
QY 301 CATATGAGAGTGTACTACCAAGGTAAACAGAAATTTGCTGCCAATTCOAATCCC 360  
|  
|  
|  
Db 19036 CATATGAGAGTGTACTACCAAGGTAAACAGAAATTTGCTGCCAATTCOAATCCC 19095  
|  
|  
|  
QY 361 TGAATGAGTGAAGAAATACGAATTAATATATGAGGGAATTCACAGAGCTGGTACGAGT 420  
|  
|  
|  
Db 19096 TGAATGAGTGAAGAAATACGAATTAATATATGAGGGAATTCACAGAGCTGGTACGAGT 19155  
|  
|  
|  
QY 421 TGCAGAGTGCAGATCTTGTCTCATCAATGACAGAGTGTGCTGTAGCAACGCTCA 480  
|  
|  
|  
Db 19156 TGCAGAGTGCAGATCTTGTCTCATCAATGACAGAGTGTGCTGTAGCAACGCTCA 19215  
|  
|  
|  
QY 481 CTGCTTCATTTCTGCTGCTGCTGCTTTAAATATGCTTTTCTCACTCAATGGCTTTCTT 540  
|  
|  
|  
Db 19216 CTGCTTCATTTCTGCTGCTGCTGCTTTAAATATGCTTTTCTCACTCAATGGCTTTCTT 19275  
|  
|  
|  
QY 541 CCTCTGCGAGTCAAGTCTTTCTTGGGTCAACAGCAATATGATCTTTAGATCACCTGGT 600  
|  
|  
|  
Db 19276 CCTCTGCGAGTCAAGTCTTTCTTGGGTCAACAGCAATATGATCTTTAGATCACCTGGT 19335  
|  
|  
|  
QY 601 ACTCAAGGAGCTACAGACATTTGGGCAATCCCACTTCCACTCTTGGAAAAACAATTTTA 660  
|  
|  
|  
Db 19336 ACTCAAGGAGCTACAGACATTTGGGCAATCCCACTTCCACTCTTGGAAAAACAATTTTA 19395  
|  
|  
|  
QY 661 TGGAAAGCAAGGTTGCCATATGCTCTTGAAGTGTGTTGCTCAGCCAAAGCCCAAGCTT 720  
|  
|  
|  
Db 19396 TGGAAAGCAAGGTTGCCATATGCTCTTGAAGTGTGTTGCTCAGCCAAAGCCCAAGCTT 19455  
|  
|  
|  
QY 721 TGTGCTTCAATGAATTAATTAAGAGCTTCAAGACAGATCCAACTTTCAATGCTCTCA 780  
|  
|  
|  
Db 19456 TGTGCTTCAATGAATTAATTAAGAGCTTCAAGACAGATCCAACTTTCAATGCTCTCA 19515  
|  
|  
|  
QY 781 CCCAATGGAATTAAGAAACAATTTGCCATATGCTCAATGCCACTTTT--TCAGGTGGGATG 839  
|  
|  
|  
Db 19516 CCCAATGGAATTAAGAAACAATTTGCCATATGCTCAATGCCACTTTTTCAGGTGGGATG 19575  
|  
|  
|  
QY 840 GTAGATGCTGGAATGGGTCAACAGACTTGCACCAACCAACTTTGCAAAAAAGGCTGGAAGC 899  
|  
|  
|  
Db 19576 GTAGATGCTGGAATGGGTCAACAGACTTGCACCAACCAACTTTGCAAAAAAGGCTGGAAGC 19635  
|  
|  
|  
QY 900 TCTGACTGGGGAACCTTAATATGCAAAAGTTATGCTCTTCAATGACGAATATGAACC 959  
|  
|  
|  
Db 19636 TCTGACTGGGGAACCTTAATATGCAAAAGTTATGCTCTTCAATGACGAATATGAACC 19695  
|  
|  
|  
QY 960 CGTGATGGAATATATCTAAAGGCTGGCTTATGTTCTATTCCTTCAAAACCTGGGTA 1019  
|  
|  
|  
Db 19696 CGTGATGGAATATATCTAAAGGCTGGCTTATGTTCTATTCCTTCAAAACCTGGGTA 19755  
|  
|  
|  
QY 1020 GAATAGATATGCTTGTCCCTTTAAAAAATGTCACAATTCATTTATGATGCTGTGTA 1079  
|  
|  
|  
Db 19756 GAATAGATATGCTTGTCCCTTTAAAAAATGTCACAATTCATTTATGATGCTGTGTA 19815  
|  
|  
|  
QY 1080 TAGTAATCAAGATCATGCTTCATGAATAATGCTTCAAGACCAATATATAGAGATTTT 1139  
|  
|  
|  
Db 19816 TAGTAATCAAGATCATGCTTCATGAATAATGCTTCAAGACCAATATATAGAGATTTT 19875  
|  
|  
|  
QY 1140 TAGCATGTTGACAAAAGAGAGGCAATTTCAAGTGTGAATTTGTTCAAGAAAGTATTTG 1199  
|  
|  
|  
Db 19876 TAGCATGTTGACAAAAGAGAGGCAATTTCAAGTGTGAATTTGTTCAAGAAAGTATTTG 19935  
|  
|  
|  
QY 1200 ATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGTGAAAACAGAGTCTCACTTTGCA 1259  
|  
|  
|  
Db 19936 ATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGTGAAAACAGAGTCTCACTTTGCA 19995  
|  
|  
|  
QY 1260 CCAGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACTCTGCTGCCAGGTTTC 1319  
|  
|  
|  
Db 19996 CCAGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACTCTGCTGCCAGGTTTC 20055  
|  
|  
|  
QY 1320 AAGGATTTCTCTGTCAGCTTCCGAATAGCTGGGATTAACAGGCGCATGCAACCAATGC 1379  
|  
|  
|

Db	20056	AAGCATCTCTCTGACGCTTCCCAATAGCTGGAGTTACAGGCCATGCAACACATGAC	20115		
Qy	1380	CTAATTTTGTATTTTGTAGTACAGAGATTTCGCAATGTGACAGGCTTGCCCTGAA	1439		
Db	20116	CTAATTTTGTATTTTGTAGTACAGAGATTTCGCAATGTGACAGGCTTGCCCTGAA	20175		
Qy	1440	CTCCTGACTTCAGGTATCCACCACTCAGCCTCCCAAGACGTGGATTACAGGCAATG	1499		
Db	20116	CTCCTGACTTCAGGTATCCACCACTCAGCCTCCCAAGACGTGGATTACAGGCAATG	20235		
Qy	1500	AGCCACCGTCCGACGCTGTTTTCAGATCCTGTA-TTTGTTCTGAAGCCTTCATTTC	1558		
Db	20236	AGCCACCGTCCGACGCTGTTTTCAGATCCTGTA-TTTGTTCTGAAGCCTTCATTTC	20295		
Qy	1559	TATCTCTTATTCATTTTGGAGTAGTACACCTAAGTAGTTTAAACAATCAATATTC	1618		
Db	20236	TATCTCTTATTCATTTTGGAGTAGTACACCTAAGTAGTTTAAACAATCAATATTC	20355		
Qy	1619	TTTGGAAAAATCCCTGCTCTTCTTATCTTACAAAAATATGTTCAAGTAGCTGATG	1678		
Db	20356	TTTGGAAAAATCCCTGCTCTTCTTATCTTACAAAAATATGTTCAAGTAGCTGATG	20415		
Qy	1679	TTATGTTTCTTCAAAATTTTCAATTTCTATCTCAGAAATTTATCTCATGCTTAATGTT	1738		
Db	20416	TTATGTTTCTTCAAAATTTTCAATTTCTATCTCAGAAATTTATCTCATGCTTAATGTT	20475		
Qy	1739	ATTGAATGCTTCGCTCTTCTGATCCAGTTTCGCTCTCTTATTTCTGCTAAAGTCTA	1798		
Db	20476	ATTGAATGCTTCGCTCTTCTGATCCAGTTTCGCTCTCTTATTTCTGCTAAAGTCTA	20535		
Qy	1799	AGTGCCTATTGAATTAAGAGCTGTGACAGATCTTTCTCCAAATATGTTATCTTTTG	1858		
Db	20536	AGTGCCTATTGAATTAAGAGCTGTGACAGATCTTTCTCCAAATATGTTATCTTTTG	20595		
Qy	1859	ACTGCATGCGACGTGACAACTGTTAAGTTTGAATTTCTTCAATTCACACAGAACT	1918		
Db	20596	ACTGCATGCGACGTGACAACTGTTAAGTTTGAATTTCTTCAATTCACACAGAACT	20655		
Qy	1919	GCTGACTCTCTCTTCCGTAAGCAATGCGCAAGCAACAGATGTTAGATATGTTAG	1978		
Db	20656	GCTGACTCTCTCTTCCGTAAGCAATGCGCAAGCAACAGATGTTAGATATGTTAG	20715		
Qy	1979	CAACAGGACATGGTGATACG 2001			
Db	20716	CAACAGGACATGGTGATACG 20738			
RESULT 4	AF222856	69058 bp	DNA	linear	PRI 17-JUL-2000
LOCUS	AF222856				Human sapiens clone Pnc chromosome 10 map 10q25.2 genomic sequence.
DEFINITION	AF222856				
ACCESSION	AF222856.1	GI:9246846			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					

/rpt\_family="Alu"  
repeat\_region 45532..45665  
/rpt\_family="MIR"  
repeat\_region complement(46257..46490)  
/rpt\_family="Alu"  
repeat\_region 48262..48578  
/rpt\_family="Alu"  
repeat\_region complement(48687..50337)  
/rpt\_family="L1"  
repeat\_region complement(50349..50611)  
/rpt\_family="Alu"  
repeat\_region complement(50670..50831)  
/rpt\_family="L1"  
repeat\_region complement(51085..51374)  
/rpt\_family="Alu"  
repeat\_region complement(52905..53197)  
/rpt\_family="Alu"  
repeat\_region 53896..54467  
/rpt\_family="LTR12"  
repeat\_region 59129..59723  
/rpt\_family="LTR12"  
repeat\_region 60018..60302  
/rpt\_family="Alu"  
repeat\_region complement(60338..60470)  
/rpt\_family="MIR"  
repeat\_region complement(61353..61521)  
/rpt\_family="MERS"  
repeat\_region 63135..63416  
/rpt\_family="Alu"  
repeat\_region 63724..63889  
/rpt\_family="MERS"  
repeat\_region 64107..64182  
/rpt\_family="MIR"  
repeat\_region 65439..65548  
/rpt\_family="MERS"  
repeat\_region 66439..66618  
/rpt\_family="MER20"  
repeat\_region complement(67562..68084)  
/rpt\_family="MER41"

## ORIGIN

Query Match 98.7%; Score 1974.2; DB 8; Length 69058;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 AATGTTGATTTCAAGGTAACAGAGCTGGAACCACTACCTGCGAGCTTGAGG 60  
DB 18658 AATGTTGATTTCAAGGTAACAGAGCTGGAACCACTACCTGCGAGCTTGAGG 18717  
QY 61 TTTCAGCGGAGCTGGTGAGCCTTGTTTTCACCTGACAGCTGGAATGAGCCAGC 120  
DB 18778 TTTCAGCGGAGCTGGTGAGCCTTGTTTTCACCTGACAGCTGGAATGAGCCAGC 18777  
QY 121 TTCAGTGAAGCTGTTCCCTCCCTCCAGGTTACCCCAATTCTCAGTCTCTCAG 180  
DB 18778 TTCAGTGAAGCTGTTCCCTCCCTCCAGGTTACCCCAATTCTCAGTCTCTCAG 18837  
QY 181 AAGCCAAAAATGAATTTGAGGGTTAGGATTTGGTCTTTTATCTATTAACAGATG 240  
DB 18838 AAGCCAAAAATGAATTTGAGGGTTAGGATTTGGTCTTTTATCTATTAACAGATG 18897  
QY 241 AATAATGTTCTCCACCGAGATGTTCTGTGTAACAATCACTCTTGACACTACG 300  
DB 18898 AATAATGTTCTCCACCGAGATGTTCTGTGTAACAATCACTCTTGACACTACG 18957  
QY 301 CATATGACAGAGTGTACTACCAAGGTAACACAGATTGGCTGCCAATTCCAATGCC 360  
DB 18958 CATATGACAGAGTGTACTACCAAGGTAACACAGATTGGCTGCCAATTCCAATGCC 19017  
QY 361 TGAATGATGAGAAATCAGAAATTAATAAGGGATTCACAGAGCTGGCTACGAGATG 420  
DB 19018 TGAATGATGAGAAATCAGAAATTAATAAGGGATTCACAGAGCTGGCTACGAGATG 19077

QY 421 TCCAGTGTGACAGATACCTTGTCTCATATAGCAGAGTGCTGCTCTACCACTGCTCA 480  
DB 19078 TCCAGTGTGACAGATACCTTGTCTCATATAGCAGAGTGCTGCTCTACCACTGCTCA 19137  
QY 481 CTGCTTCATTTCCCTGCTTGCTGCTTTAAATCTGCTTTTCTCAGCTCAATTGGCTTCTT 540  
DB 19138 CTGCTTCATTTCCCTGCTTGCTGCTTTAAATCTGCTTTTCTCAGCTCAATTGGCTTCTT 19197  
QY 541 CCCTGTGAGCAGTACAGTTTCTTTGGGTCAACAGCAAAAGATCTTTTAAATCAACCTGGT 600  
DB 19198 CCCTGTGAGCAGTACAGTTTCTTTGGGTCAACAGCAAAAGATCTTTTAAATCAACCTGGT 19257  
QY 601 ACTCAAGAGCTACAGACATTTGGGCATCCACTTCACTCTCTTGGAAAAACAATTTTA 660  
DB 19258 ACTCAAGAGCTACAGACATTTGGGCATCCACTTCACTCTCTTGGAAAAACAATTTTA 19317  
QY 661 TGAAGCCAGAGTTGCATATGTCCTCTTGAAGTTGTTTGTCAAGCCAGCCCAAGCTT 720  
DB 19318 TGAAGCCAGAGTTGCATATGTCCTCTTGAAGTTGTTTGTCAAGCCAGCCCAAGCTT 19377  
QY 721 TGTGCTTCAACATGAATTAAGAGCTTCAACAGATTCACATTTTCAATGCGCTCA 780  
DB 19378 TGTGCTTCAACATGAATTAAGAGCTTCAACAGATTCACATTTTCAATGCGCTCA 19437  
QY 781 CCCAATGGAATAAAGAACAAATGGCATATCTCAATGACCACTTTT-TCAGTGGAGATG 839  
DB 19438 CCCAATGGAATAAAGAACAAATGGCATATCTCAATGACCACTTTTCTCAGTGGAGATG 19497  
QY 840 GTAGATGCTGAATGGGTCAAGCATTTGCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 899  
DB 19498 GTAGATGCTGAATGGGTCAAGCATTTGCCAACCAAACTTTGCAAAAAAGGCTGGAAGC 19557  
QY 900 TCTGACTGGGAGCCTTAATATGCAAAAGTTATAGGCTCTTATGACAGAAATATGAACC 959  
DB 19558 TCTGACTGGGAGCCTTAATATGCAAAAGTTATAGGCTCTTATGACAGAAATATGAACC 19617  
QY 960 CGTATATGATATATAGCTAAAGGTTGGCCTTATGTTCTATTCCTTCAAAACCTGGTA 1019  
DB 19618 CGTATATGATATATAGCTAAAGGTTGGCCTTATGTTCTATTCCTTCAAAACCTGGTA 19677  
QY 1020 GAATGATATGCTGTTTCCCTTTAAAAATGTCAAAATTCGATTTATGATGCTGTGTA 1079  
DB 19678 GAATGATATGCTGTTTCCCTTTAAAAATGTCAAAATTCGATTTATGATGCTGTGTA 19737  
QY 1080 TAGTAATCAAGATCATGCTCATGAAATGCTTCAGAACCCATATATAGAGATTTT 1139  
DB 19738 TAGTAATCAAGATCATGCTCATGAAATGCTTCAGAACCCATATATAGAGATTTT 19797  
QY 1140 TAGCATGTGTACAAAAGAGAGGCATTTCACTGTTGAATTTGTTACAGAGAATTTG 1199  
DB 19798 TAGCATGTGTACAAAAGAGAGGCATTTCACTGTTGAATTTGTTACAGAGAATTTG 19857  
QY 1200 AATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTCA 1259  
DB 19858 AATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTCA 19917  
QY 1260 CCCAGGCTGAGATACAGTGGCTGTGTGCTCGGCTCACTGCAACCTGTGCTCCAGGTTTC 1319  
DB 19918 CCCAGGCTGAGATACAGTGGCTGTGTGCTCGGCTCACTGCAACCTGTGCTCCAGGTTTC 19977  
QY 1320 AAGGATTTCTGCTGACGTTCCGAAATAGCTGGGATTAACAGGGGATGCAACCAATGC 1379  
DB 19978 AAGGATTTCTGCTGACGTTCCGAAATAGCTGGGATTAACAGGGGATGCAACCAATGC 20037  
QY 1380 CTAATTTTGTATTTTATAGTAGACAGATTTGCGCATGTTGACCAAGCTTGCCTTGA 1439  
DB 20038 CTAATTTTGTATTTTATAGTAGACAGATTTGCGCATGTTGACCAAGCTTGCCTTGA 20097  
QY 1440 CTCCTGACTGACAGGATCCACCCGACCTCAGGCTCCAAAGACAGGAGATTAACAGGATG 1499  
DB 20098 CTCCTGACTGACAGGATCCACCCGACCTCAGGCTCCAAAGACAGGAGATTAACAGGATG 20157  
QY 1500 AGCCACCGTGCCAGCCTGTTTCTCAGATCTGTA-TTGTGTTCTGAAGCCTTCATTTTC 1558

Db 20158 AGCCACCGGCCACCGCTGTTTTCAGATCTCGATTTTGTTCGAGCCCTCATTTTC 20217  
Oy 1559 TATCTTCTTAATCATTTTGGAGTAGTACACCTAAGTAAGTTTTTAAACAATCAATATTC 1618  
Db 20218 TATCTTCTTAATCATTTTGGAGTAGTACACCTAAGTAAGTTTTTAAACAATCAATATTC 20217  
Oy 1619 TTTGGAAAAATTCCTGGTTCCTTTCTTAATCTCTTACAAAAAATATGTTCAAGTATAGCTGATG 1678  
Db 20278 TTTGGAAAAATTCCTGGTTCCTTTCTTAATCTCTTACAAAAAATATGTTCAAGTATAGCTGATG 20337  
Oy 1679 TTAATGTTCTTCAATATTAATTTCTCTATCTCAGAAATTTATCTCATGCTTAATGTT 1738  
Db 20338 TTAATGTTCTTCAATATTAATTTCTCTATCTCAGAAATTTATCTCATGCTTAATGTT 20397  
Oy 1739 ATTGAATAGTCTTCACTCTTCTTCATCCAGTTTCGTCTCTTAATTTCACTCAAGTCTTA 1798  
Db 20398 ATTGAATAGTCTTCACTCTTCTTCATCCAGTTTCGTCTCTTAATTTCACTCAAGTCTTA 20457  
Oy 1799 AGTGGCTATTAAGATTAAGAGCTTTGATACAGATTTCTTCTCAATATGTTATCTTTTG 1858  
Db 20458 ATTGCTATTAAGATTAAGAGCTTTGATACAGATTTCTTCTCAATATGTTATCTTTTG 20517  
Oy 1859 ACTGCATGCGAATGACAACTGTAACTGTTTGATTTCTCAATATCCACAGAACT 1918  
Db 20518 ACTGCATGCGAATGACAACTGTAACTGTTTGATTTCTCAATATCCACAGAACT 20577  
Oy 1919 GCTGACTCTCTCTTCTCTTGAAGCAATGCCAAGCACACAGTATGTATATGTATAG 1978  
Db 20578 GCTGACTCTCTCTTCTCTTGAAGCAATGCCAAGCACACAGTATGTATATGTATAG 20637  
Oy 1979 CAACAGGACATGGGTGTCATAGC 2001  
Db 20638 CAACAGGACATGGGTGTCATAGC 20660

RESULT 5  
AF042484 80155 bp DNA linear PRI 22-JUN-2000  
LOCUS Homo sapiens sequence of the 80 kb core centromere protein binding  
DEFINITION region of the mar del(10) neocentromere corresponding to human  
chromosome 10q25.2, complete sequence.  
ACCESSION AF042484  
VERSION AF042484.1 GI:4205782  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 80155)  
AUTHORS Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.  
TITLE Sequence analysis of an 80 kb human neocentromere  
JOURNAL Hum. Mol. Genet. 8 (2), 217-227 (1999)  
PUBMED 9931329  
2 (bases 1 to 80155)  
REFERENCE 2 (bases 1 to 80155)  
AUTHORS Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,A.  
TITLE Direct Subdivision  
JOURNAL Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch  
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,  
Parkville, Melbourne, Victoria 3052, Australia  
FEATURES  
source  
1..80155  
location/Qualifiers  
1..80155  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10q25.2"  
/note="Sequence information obtained from BAC clones  
prepared from cell line BE2C1-18-5F"  
repeat\_region  
complement(join(49..525,3628..4545,4849..5053))  
/rpt\_family="LIMEC"  
/rpt\_type=dispersed

repeat\_region join(560..597,3184..3281)  
/rpt\_family="MER58"  
/rpt\_type=dispersed  
repeat\_region complement(join(598..2553,2716..3183))  
/rpt\_family="TIGGER1"  
/rpt\_type=dispersed  
repeat\_region 2558..2581  
/standard\_name="T1"  
/note="100% homology"  
/rpt\_type=direct  
repeat\_region 2570..2715  
/rpt\_family="AluSg/x"  
/rpt\_type=dispersed  
repeat\_region 3021..3050  
/standard\_name="T2"  
/note="83.3% homology"  
/rpt\_type=direct  
repeat\_region 3283..3558  
/rpt\_family="AluSx"  
/rpt\_type=dispersed  
repeat\_region 4173..4211  
/standard\_name="T3"  
/note="82.1% homology"  
/rpt\_type=direct  
repeat\_region complement(4546..4848)  
/rpt\_family="AluSx"  
/rpt\_type=dispersed  
repeat\_region complement(5909..6208)  
/rpt\_family="AluSg"  
/rpt\_type=dispersed  
repeat\_region join(6232..6573,6638..6688)  
/rpt\_family="MLT2CB"  
/rpt\_type=dispersed  
repeat\_region 6598..6645  
/standard\_name="AC24"  
/note="100% homology"  
/rpt\_type=tandem  
repeat\_region 6598..6637  
/note="microsatellite"  
/rpt\_type=tandem  
repeat\_region 7091..7134  
/rpt\_unit="ca"  
repeat\_region 7808..7911  
/rpt\_family="MIR"  
/rpt\_type=dispersed  
repeat\_region 8349..8524  
/rpt\_type=dispersed  
repeat\_region 8349..8524  
/rpt\_family="MIR"  
/rpt\_type=dispersed  
repeat\_region complement(8839..9124)  
/rpt\_family="AluJb"  
/rpt\_type=dispersed  
repeat\_region 9180..9280  
/rpt\_family="MIR"  
repeat\_region 10070..10170  
/rpt\_type=dispersed  
repeat\_region 10097..10136  
/note="low complexity"  
/rpt\_family="purine-rich"  
repeat\_region 10175..10339  
/standard\_name="GA1"  
/note="82.5% homology"  
/rpt\_type=direct  
repeat\_region 10175..10339  
/rpt\_family="PRAM"  
/rpt\_type=dispersed  
repeat\_region complement(10899..11197)  
/rpt\_family="AluSx"

```
misc_feature
join(11627..11757,12434..12512,13443..13562)
/note="98% identity"
/db_xref="dbEST:AA774571"
repeat_region
/standard_name="AT4"
/note="100% homology"
/rpt_type=tandem
/rpt_unit="(attc)8"
repeat_region
/complement(12010..12312)
/rpt_family="AluDb"
/rpt_type=dispersed
12327..12433
/rpt_family="L2"
/rpt_type=dispersed
complement(13946..14235)
/rpt_family="AluSg"
/rpt_type=dispersed
15130..15159
/note="low complexity"
/rpt_family="AT-rich"
15164..15753
/standard_name="AT28"
/note="82.1% homology"
/rpt_type=direct
/rpt_unit="(atgatatatgctatagacataaac)21"
join(15170..15289,15410..15612)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="ca"
15617..15731
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="cata"
16004..16100
/rpt_family="L2"
/rpt_type=dispersed
complement(16128..16364)
/rpt_family="L2"
repeat_region
/rpt_type=dispersed
complement(16407..16592)
/rpt_family="L2"
repeat_region
/rpt_type=dispersed
complement(16657..16889)
/rpt_family="MIR"
/rpt_type=dispersed
16914..16936
/note="low complexity"
/rpt_family="AT-rich"
17102..17386
/rpt_family="AluSx"
/rpt_type=dispersed
17406..17564
/standard_name="GA2"
/note="81.8% homology"
/rpt_type=direct
17416..17543
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="gaaa"
17647..17702
/rpt_family="MIR"
/rpt_type=dispersed
complement(19863..20161)
/rpt_family="AluSx"
/rpt_type=dispersed
complement(20842..20913)
/rpt_family="L2"
/rpt_type=dispersed
complement(21366..21414)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="gaaa"
```

```
repeat_region 21375..21639
/standard_name="CT1"
/note="75% homology"
/rpt_type=dispersed
repeat_region
/complement(21415..21536)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="cgaaa"
repeat_region
/complement(21524..21646)
/note="low complexity"
/rpt_family="purine-rich"
22409..22646
/rpt_family="L1M4"
/rpt_type=dispersed
repeat_region
22560..22604
/standard_name="T4"
/note="80% homology"
/rpt_type=direct
/rpt_unit="(aatattacataatc)3"

Query Match 98.7%; Score 1974.2; DB 8; Length 80155;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 AATGTTGGATTTCAAAGGTAAACAAGCTGGAACCAACTACGCGCGAGCTTGAGAG 60
DB 18645 AATGTTGGATTTCAAAGGTAAACAAGCTGGAACCAACTACGCGCGAGCTTGAGAG 18704

QY 61 TTTCAGCGGAGACTGGTGCAGGCTTGTGTTTTCACCTGACAGCTGAAATGAGCCAGC 120
DB 18705 TTTCAGCGGAGACTGGTGCAGGCTTGTGTTTTCACCTGACAGCTGAAATGAGCCAGC 18764

QY 121 TTCAGTGAAGCTGTTGTTTCTCCCTCCCAAGSTTACCAGCAATTCTGACTTCTCAG 180
DB 18765 TTCAGTGAAGCTGTTGTTTCTCCCTCCCAAGSTTACCAGCAATTCTGACTTCTCAG 18824

QY 181 AAAGCAAAATGAATTTAGAGGTTTGGATTGGTTCTTTATATTAACAGATG 240
DB 18825 AAAGCAAAATGAATTTAGAGGTTTGGATTGGTTCTTTATATTAACAGATG 18884

QY 241 ATATATATGTTCTCCACAGATGTTCTGTTGAACAATCACTCACTTCCGACACTACG 300
DB 18885 ATATATATGTTCTCCACAGATGTTCTGTTGAACAATCACTCACTTCCGACACTACG 18944

QY 301 CATATGCGAGAGTGTACTACCAAGTTAAACAAGAAATTGGCTGCCAATTCGAATCCC 360
DB 18945 CATATGCGAGAGTGTACTACCAAGTTAAACAAGAAATTGGCTGCCAATTCGAATCCC 19004

QY 361 TGAATCGATGAGAAATTCGAATTATATATGGGGAATTCAACAGAGCTGGCTAGGAG 420
DB 19005 TGAATCGATGAGAAATTCGAATTATATATGGGGAATTCAACAGAGCTGGCTAGGAG 19064

QY 421 TGCAGTGTGTCAGATTAATCTTGTCAATACAGAGTGTGCTGCTTACGAATGCTCA 480
DB 19065 TGCAGTGTGTCAGATTAATCTTGTCAATACAGAGTGTGCTGCTTACGAATGCTCA 19124

QY 481 CTGCTTCAATTCCTGCTTGGTCTTTAAATACGCTTTTCTCAAGTCAATTGGCTTTCT 540
DB 19125 CTGCTTCAATTCCTGCTTGGTCTTTAAATACGCTTTTCTCAAGTCAATTGGCTTTCT 19184

QY 541 CCTCTGCGAGTCAGATTTCTTGGGTCAACAGAAATGATCTTTGAAATCACCTGGT 600
DB 19185 CCTCTGCGAGTCAGATTTCTTGGGTCAACAGAAATGATCTTTGAAATCACCTGGT 19244

QY 601 ACTCAAAAGAGCTACAAACATTTGGGCATTCACCTTCCACTCTTGGGAAAAACAATTTTA 660
DB 19245 ACTCAAAAGAGCTACAAACATTTGGGCATTCACCTTCCACTCTTGGGAAAAACAATTTTA 19304

QY 661 TGGAAAGCCAGGTTCCATAGTGCTCTTGAAGTTGTTTGTCAAGCCAAAGCCCAAGCTT 720
DB 19305 TGGAAAGCCAGGTTCCATAGTGCTCTTGAAGTTGTTTGTCAAGCCAAAGCCCAAGCTT 19364

QY 721 TGTGCTTCAAAATGAATTAAGAGCTTCAGAAACAAGATCCAAATTTTCATAGGCTCA 780
```

Db	19365	TTGCTTCAACATGAAATTTAGAGCTTCAGAACAGATCCAAATTTTCAATTTGCTCTCA	19424
QY	781	CCCAACTGATTAATAAGAAACATTGCCATATCTCAATGACCCTTT-TCAGGTGGAG	839
Db	19425	CCCAACTGATTAATAAGAAACATTGCCATATCTCAATGACCACCTTTTCTCAGGTGGAG	19484
QY	840	GTAGATCTGTGATGGGTCAACAGATTGCCCAACAACTTTGCCAAAAAGCTGGAGC	899
Db	19485	GTAGATCTGTGAAATGGGTCAACAGATTGCCCAACAACTTTGCCAAAAAGCTGGAGC	19544
QY	900	TCGACGTGGGGACCCCTAAATATGCAAAAGTTGATAGGCTCTTCATGACAGATATGAAACC	959
Db	19545	TCGACGTGGGGACCCCTAAATATGCAAAAGTTAATAGGCTCTTCATGACAGATATGAAACC	19604
QY	960	CGTGTATGATATATAGCTAAAGGGTTGGCCCTTTATGTTTCATTCCTTCAAAACCTGGTA	1019
Db	19605	CGTGTATGATATATAGCTAAAGGGTTGGCCCTTTATGTTTCATTCCTTCAAAACCTGGTA	19664
QY	1020	GAATATGATATGCTGTTTCCCTTTAAAAAATGTCACAAATGCAATTTATGATGCTGTGA	1079
Db	19665	GAATATGATATGCTGTTTCCCTTTAAAAAATGTCACAAATGCAATTTATGATGCTGTGA	19724
QY	1080	TAGTATCTCAAGATCAATGCTCCATGGAATNGCTGACGAACCAATATATAGAGATTTT	1139
Db	19725	TAGTATCTCAAGATCAATGCTCCATGGAATNGCTGACGAACCAATATATAGAGATTTT	19784
QY	1140	TAGCATGTGTGACAAAAAGAGAGGCCATTTCAGTGTGAATGTTCAAGAGAATTTG	1199
Db	19785	TAGCATGTGTGACAAAAAGAGAGGCCATTTCAGTGTGAATGTTCAAGAGAATTTG	19844
QY	1200	ATATATGTTTTCTCAGATCTTTTATTTTTTATTTTTTTGAAACAGAGTCTCACTTTGTCA	1259
Db	19845	ATATATGTTTTCTCAGATCTTTTATTTTTTATTTTTTTGAAACAGAGTCTCACTTTGTCA	19904
QY	1260	CCGAGCTGAGATACATGCGCTGTGTCTGGGCTCATGCAACTGCGCTCCAGGTTCC	1319
Db	19905	CCGAGCTGAGATACATGCGCTGTGTCTGGGCTCATGCAACTGCGCTCCAGGTTCC	19964
QY	1320	AAGGATTTCTCTGTCAAGCTCCCGAATATGCTGGATTTACAGGCGCATGACCAACATGC	1379
Db	19965	AAGGATTTCTCTGTCAAGCTCCCGAATATGCTGGATTTACAGGCGCATGACCAACATGC	20024
QY	1380	CTAATTTTTGTATTTTTTAGTAGACAGAGTTTCGCATGTTGAACAGAGCTTGCCTTGA	1439
Db	20025	CTAATTTTTGTATTTTTTAGTAGACAGAGTTTCGCATGTTGAACAGAGCTTGCCTTGA	20084
QY	1440	CTCTCGATTTAGGTGATCCACCACCTCAGCTCCCAAGCACTGGGATTAAGGCATG	1499
Db	20085	CTCTCGATTTAGGTGATCCACCACCTCAGCTCCCAAGCACTGGGATTAAGGCATG	20144
QY	1500	AGCCACGTCGCCAGCGCTGTTTTCTCAGATCCCTGTA-TTTGTTTCTGAAGCTTCAATTC	1558
Db	20145	AGCCACGTCGCCAGCGCTGTTTTCTCAGATCCCTGTA-TTTGTTTCTGAAGCTTCAATTC	20204
QY	1559	TATCTCTTAATCAATTTTGAAGTAGTACACTAATAGGTTTTTTAACAATCAATATTC	1618
Db	20205	TATCTCTTAATCAATTTTGAAGTAGTACACTAATAGGTTTTTTAACAATCAATATTC	20264
QY	1619	TTTGGAAATTCCTGTTCTTTCTTATTTCTTACAAAATATGTTTCAATATAGCTGATG	1678
Db	20265	TTTGGAAATTCCTGTTCTTTCTTATTTCTTACAAAATATGTTTCAATATAGCTGATG	20324
QY	1679	TTATGTTCTTCAAAATATCATTTTCATATCAGAAATTTATATGCTCAATATGTT	1738
Db	20325	TTATGTTCTTCAAAATATCATTTTCATATCAGAAATTTATATGCTCAATATGTT	20384
QY	1739	ATTGAATAGTCTCACTTCTGTGATCAGATTTCTGGTCTTATTTCACTTAAGTCTTA	1798
Db	20385	ATTGAATAGTCTCACTTCTGTGATCAGATTTCTGGTCTTATTTCACTTAAGTCTTA	20444
QY	1799	AGTGGCTATTAAGATTAAGAGCTTGTAACAGATTTCTTTCGAATATGCTTATCTTTTG	1858

Db	20445	ATTGGCGATTATGAAATAAAGAGCTTTGTAACAGATTCTTTCGCCAATATGCTTATCTTTTG	20504
QY	1859	ACTGCATGCCAGTGACAACTGTTTAACTGTTTGATTTCTTCATPAACATTCACAGAACT	1918
Db	20505	ACTTCATGACGAGTGAACAACTGTTAACTGTTTGATTTCTTCATPAACATTCACAGAACT	20564
QY	1919	GCTACATCCTCTCTCTCTGTAAGCAAGTGGCCAGGACACAGCATTTGTAATAGTATGTAG	1978
Db	20565	GCTACATCCTCTCTCTCTGTAAGCAAGTGGCCAGGACACAGCATTTGTAATAGTATGTAG	20624
QY	1979	CAACAGGAGACATGGGTGCATAGC	2001
Db	20625	CAACAGGAGACATGGGTGCATAGC	20647
RESULT 6			
LOCUS	AF222854	80202 bp	DNA linear PRI 17-JUL-2000
DEFINITION	Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.		
ACCESSION	AF222854		
VERSION	AF222854.1	GI:9246844	
KEYWORDS	.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 80202) Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M., Irvine,D.V., Saffery,R. and Choo,K.H.		
TITLE	The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification		
JOURNAL	Genome Res. 10 (6), 832-838 (2000)		
PUBMED	10854414		
REFERENCE	2 (bases 1 to 80202)		
AUTHORS	Barry,A.E.		
JOURNAL	Direct Submission Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville, Melbourne, Victoria 3052, Australia		
REMARK	Human genomic sequence from 10q25.2, clone1b=NC, second release		
FEATURES	Location/Qualifiers		
SOURCE	1..80202		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="10"		
	/map="10q25.2"		
	/clone="NC"		
	complement(598..3183)		
	/rpt_family="11ggr1"		
	2587..2715		
	/rpt_family="Alu"		
	3283..3553		
	/rpt_family="Alu"		
	complement(4548..4809)		
	/rpt_family="Alu"		
	complement(5919..6211)		
	/rpt_family="Alu"		
	6257..6599		
	/rpt_family="MLT2B2"		
	complement(8848..9123)		
	/rpt_family="Alu"		
	complement(10913..11198)		
	/rpt_family="Alu"		
	complement(12062..12313)		
	/rpt_family="Alu"		
	complement(13957..14236)		
	/rpt_family="Alu"		
	17133..17424		
	/rpt_family="Alu"		
	complement(19882..20197)		

```
repeat_region /rpt_family="Alu" complement(22664. .22956)
repeat_region /rpt_family="Alu" complement(23935. .24113)
repeat_region /rpt_family="MIR" complement(24884. .24884)
repeat_region /rpt_family="Alu" complement(25618. .25973)
repeat_region /rpt_family="THE1" complement(26970. .26970)
repeat_region /rpt_family="MER2" complement(27449. .27449)
repeat_region /rpt_family="Alu" complement(27996. .27996)
repeat_region /rpt_family="Alu" complement(28978. .28978)
repeat_region /rpt_family="Alu" complement(30668. .30736)
repeat_region /rpt_family="MLT1" complement(31460. .31993)
repeat_region /rpt_family="MER1" complement(32130. .32248)
repeat_region /rpt_family="MLT1" complement(32961. .32961)
repeat_region /rpt_family="Alu" complement(33403. .33403)
repeat_region /rpt_family="Alu" complement(34169. .34233)
repeat_region /rpt_family="MIR" complement(34639. .34639)
repeat_region /rpt_family="Alu" complement(35993. .36081)
repeat_region /rpt_family="MIR" complement(38438. .38721)
repeat_region /rpt_family="Alu" complement(39333. .39333)
repeat_region /rpt_family="Alu" complement(42817. .42877)
repeat_region /rpt_family="MIR" complement(44845. .45119)
repeat_region /rpt_family="Alu" complement(45665. .45665)
repeat_region /rpt_family="MIR" complement(46257. .46490)
repeat_region /rpt_family="Alu" complement(48578. .48578)
repeat_region /rpt_family="Alu" complement(48687. .50337)
repeat_region /rpt_family="L1" complement(50349. .50611)
repeat_region /rpt_family="Alu" complement(50670. .50831)
repeat_region /rpt_family="L1" complement(51085. .51374)
repeat_region /rpt_family="Alu" complement(52905. .53197)
repeat_region /rpt_family="Alu" complement(54467. .54467)
repeat_region /rpt_family="LTR12" complement(59129. .59719)
repeat_region /rpt_family="LTR12" complement(60302. .60302)
repeat_region /rpt_family="Alu" complement(60338. .60470)
repeat_region /rpt_family="MIR" complement(61353. .61521)
repeat_region /rpt_family="MERS" complement(63416. .63416)
repeat_region /rpt_family="Alu" complement(63724. .63889)
repeat_region /rpt_family="MERS" complement(64107. .64182)
repeat_region /rpt_family="MIR"
```

```
repeat_region 65459. .65548
repeat_region /rpt_family="MERS"
repeat_region 66430. .66618
repeat_region /rpt_family="MER20"
repeat_region complement(67562. .68084)
repeat_region /rpt_family="MER41"
repeat_region complement(72861. .73139)
repeat_region /rpt_family="Alu"
repeat_region 73756. .74037
repeat_region /rpt_family="Alu"
repeat_region 75588. .75853
repeat_region /rpt_family="Alu"
repeat_region complement(77441. .77729)
repeat_region /rpt_family="Alu"

ORIGIN
Query Match 98.7%; Score 1974.2; DB 8; Length 80202;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 AATGTTGGATTTCAAAGTAACAAGAGCTGGAAAACAATCTGCGCCGAGCTTGAGAG 60
Db 18658 AATGTTGGATTTCAAAGTAACAAGAGCTGGAAAACAATCTGCGCCGAGCTTGAGAG 18717
QY 61 TTTCAGCGGAGACTGCTGAGCTGAGCTTGTGTTTTCAGTGAAGAGTGAATGAGCCAGC 120
Db 18718 TTTCAGCGGAGACTGCTGAGCTGAGCTTGTGTTTTCAGTGAAGAGTGAATGAGCCAGC 18777
QY 121 TTCAGTGAAGCTTGTTCCTTCCTCCCTCAAGGTACCCACAAATTCAGTTCCTCAGG 180
Db 18778 TTCAGTGAAGCTTGTTCCTTCCTCCCTCAAGGTACCCACAAATTCAGTTCCTCAGG 18837
QY 161 AAAGCCAAAATAAGATTGAGGCTTTAGAGTTGAGTTGGTTCTTTATCTATTAACAGATTG 240
Db 18838 AAAGCCAAAATAAGATTGAGGCTTTAGAGTTGAGTTGGTTCTTTATCTATTAACAGATTG 18897
QY 241 ATATATATGTCCTCCACAGAGTTCTGCTGTATACATATCACTTCCTGACATACG 300
Db 18988 ATATATATGTCCTCCACAGAGTTCTGCTGTATACATATCACTTCCTGACATACG 18957
QY 301 CATATGAGAGTGTATTACCAAGGTAAACAGAAATGGCTGCCAATTCCAATCCC 360
Db 18958 CATATGAGAGTGTATTACCAAGGTAAACAGAAATGGCTGCCAATTCCAATCCC 19017
QY 361 TGAATCTGATGAGAGAAATCAGAAATTAATATGAGGATTTCAACAGAGCTGATGAGATG 420
Db 19018 TGAATCTGATGAGAGAAATCAGAAATTAATATGAGGATTTCAACAGAGCTGATGAGATG 19077
QY 421 TGCCAGGTGATGATCTTGTCTCATCAGAGAGTGTGCTGTAGCAATGCTCA 480
Db 19078 TGCCAGGTGATGATCTTGTCTCATCAGAGAGTGTGCTGTAGCAATGCTCA 19137
QY 481 CTGCTTCATTTCTGCTTGGTGTCTTAAATATCTGTTTTCAGCTCAATGGCTTTCTT 540
Db 19138 CTGCTTCATTTCTGCTTGGTGTCTTAAATATCTGTTTTCAGCTCAATGGCTTTCTT 19197
QY 541 CCTCTGGAGTCAAGTTCCTTGGGTCAACAGCAAAATGATCTTTAGATCACTGCT 600
Db 19198 CCTCTGGAGTCAAGTTCCTTGGGTCAACAGCAAAATGATCTTTAGATCACTGCT 19257
QY 601 ACTCAAGAGCTACAAGCAATGGGCATCCACTTCCTGGAAAAACAATTTTA 660
Db 19258 ACTCAAGAGCTACAAGCAATGGGCATCCACTTCCTGGAAAAACAATTTTA 19317
QY 661 TGGAGCCAGGTTGCCATAGTCTCTTGGAGTGTGTTGCTCAGCCAGGCCAAGCTT 720
Db 19318 TGGAGCCAGGTTGCCATAGTCTCTTGGAGTGTGTTGCTCAGCCAGGCCAAGCTT 19377
QY 721 TGTGCTTCAACATGAATTAAGAGCTTCAGAGCAAGATCCACTTTTCATATGCGCTCA 780
Db 19378 TGTGCTTCAACATGAATTAAGAGCTTCAGAGCAAGATCCACTTTTCATATGCGCTCA 19437
QY 781 CCCAATGATTAAGAAACAATGGCATATCTCATATGACCACTTTT-TCAAGTGGATG 839
```



```

|||||
Db      19438  CCGAATGATTAAGAAACAATTGCGATATCTCAATGACCACTTTTCTCAGAGTGGAGT 19497
|||||
Qy      840    GTAGATGCTGGAAATGGGTCAACAGCATTTGCCAACCAACTTTGGCAAAAAAGGCTGAAAG 899
|||||
Db      19498  GTAGATGCTGGAAATGGGTCAACAGCATTTGCCAACCAACTTTGGCAAAAAAGGCTGAAAG 19557
|||||
Qy      900    TCTGACTGGGGACCTTAATATGACAAAGTTGATAGGCTCTTCATGACGAATATGAAACC 959
|||||
Db      19558  TCTGACTGGGGACCTTAATATGACAAAGTTGATAGGCTCTTCATGACGAATATGAAACC 19617
|||||
Qy      960    CGTGTATGATATACCTAAAGGTTGGCTTTATGTTTCTATTCCTTCACAAACCTGGTA 1019
|||||
Db      19618  CGTGTATGATATACCTAAAGGTTGGCTTTATGTTTCTATTCCTTCACAAACCTGGTA 19677
|||||
Qy      1020    GAATGATATGCTGTTTCCCTTTAAAAAAGTCAACATTTGATATGATGCTGTGTA 1079
|||||
Db      19678  GAATGATATGCTGTTTCCCTTTAAAAAAGTCAACATTTGATATGATGCTGTGTA 19737
|||||
Qy      1080    TAGTAATCTCACAGATCATGCTCCATGAAATGCTTCAGAACCCAAATATAGAGATTTT 1139
|||||
Db      19738  TAGTAATCTCACAGATCATGCTCCATGAAATGCTTCAGAACCCAAATATAGAGATTTT 19797
|||||
Qy      1140    TAGCATGTGTGACAAAGAGGSCATTTCACTGTGTAATTTGTTCAAGAGAATTTTG 1199
|||||
Db      19798  TAGCATGTGTGACAAAGAGGSCATTTCACTGTGTAATTTGTTCAAGAGAATTTTG 19857
|||||
Qy      1200    ATTATGTTTCTCAGATCTTTTATTTTATTTTAAAAAGAGTGTCACTTTGTGA 1259
|||||
Db      19858  ATTATGTTTCTCAGATCTTTTATTTTATTTTAAAAAGAGTGTCACTTTGTGA 19917
|||||
Qy      1260    CCCAGGCTGAGATGACAGTGGCTGTGGTCTCAGCTCACTGCAACCTTCGCTCCAGGTT 1319
|||||
Db      19918  CCCAGGCTGAGATGACAGTGGCTGTGGTCTCAGCTCACTGCAACCTTCGCTCCAGGTT 19977
|||||
Qy      1320    AAGCATTTCTCTGTCAGCTTCCCGAATGCTGGGATTACAGGCGCATGCACCACTATC 1379
|||||
Db      19978  AAGCATTTCTCTGTCAGCTTCCCGAATGCTGGGATTACAGGCGCATGCACCACTATC 20037
|||||
Qy      1380    CPAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCGATTTGACCGAGCTTGCCCTGAA 1439
|||||
Db      20038  CPAATTTTTGTATTTTTAGTAGAGACAGAGTTTCGCGATTTGACCGAGCTTGCCCTGAA 20097
|||||
Qy      1440    CTCCTGACTTCAGTGATCCACCACTGAGCTCCCAAGGACTGGGATTACAGGCATG 1499
|||||
Db      20098  CTCCTGACTTCAGTGATCCACCACTGAGCTCCCAAGGACTGGGATTACAGGCATG 20157
|||||
Qy      1500    AGCCACCGTGCCGAGCTGTTTCTCAGATTCCTGTA-TTTGTTTCTGAAAGCTTCATTT 1558
|||||
Db      20158  AGCCACCGTGCCGAGCTGTTTCTCAGATTCCTGTA-TTTGTTTCTGAAAGCTTCATTT 20217
|||||
Qy      1559  TATCTTCTTATCATTTTGAAGTATGACACCTAAGTATTTTAAACAAATATATC 1618
|||||
Db      20218  TATCTTCTTATCATTTTGAAGTATGACACCTAAGTATTTTAAACAAATATATC 20277
|||||
Qy      1619  TTTGAAAAATTCCTGCTCTTTCTTATTCCTAACAATAATGTTCACTATAGCTGATG 1678
|||||
Db      20278  TTTGAAAAATTCCTGCTCTTTCTTATTCCTAACAATAATGTTCACTATAGCTGATG 20337
|||||
Qy      1679  TTAGTTTCTTCAAAATATATCATTTCTATCTCAGAAATTAATCTCATGCTTAATGTT 1738
|||||
Db      20338  TTAGTTTCTTCAAAATATATCATTTCTATCTCAGAAATTAATCTCATGCTTAATGTT 20397
|||||
Qy      1739  ATTGAATATGCTTCATCTTGTGCAATCCAGTTTCTGTCTCTATTTTCACTTAAGTCA 1798
|||||
Db      20398  ATTGAATATGCTTCATCTTGTGCAATCCAGTTTCTGTCTCTATTTTCACTTAAGTCA 20457
|||||
Qy      1799  AATGAGTATTAAGATAAAGAGCTGTGAACAGATTTCTTCCCAATATGATCTTCTTTG 1858
|||||
Db      20458  AATGAGTATTAAGATAAAGAGCTGTGAACAGATTTCTTCCCAATATGATCTTCTTTG 20517
|||||
Qy      1859  ACTGATGCGCAGTACAAACCTGTTAACTGTTTGAATCTTCATTAACATTCACAGAACT 1918
|||||

```

```

Db      20518  ACTGATGCCAGTGAACAAACTGTTACTGTTTGTGATTTCTTATTAACATTTCCACAGAACAT 20577
Qy      1919  GCTGACTCTCTCTCTTCTCTGAAAGCAATGCCACAGACAGCATTTGTATGATATGTAAG 1978
Db      20578  GCTGACTCTCTCTCTTCTCTGAAAGCAATGCCACAGACAGCATTTGTATGATATGTAAG 20637
Qy      1979  CAACAGGACATGGGTGATAGC 2001
Db      20638  CAACAGGACATGGGTGATAGC 20660

RESULT 7
AL355340
LOCUS
DEFINITION
  Human DNA sequence from clone Rpl1-383C6 on chromosome 10 Contains
  the TRUB1 gene for Trub pseudouridine (psi) synthase homolog 1 (E.
  coli), a novel gene and a Cpg island, complete sequence.
ACCESSION
  AL355340
VERSION
  AL355340.17 GI:21436504
KEYWORDS
  HTG; TRUB1.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Homnidae; Homo.
  1 (bases 1 to 112442)
REFERENCE
  AUTHORS
  Leonamornlert,D.
  TITLE
  Direct Submission
  JOURNAL
  Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
  Clome request@sanger.ac.uk
  On Jun 17, 2002 this sequence version replaced gi:16416169.
  The following abbreviations are used to associate primary accession
  numbers given in the feature table with their source databases:
  Em; EMBL; SW; SWISSPROT; Tr; TREMBL; Mp; WORMPEP; Information
  on the WORMPEP database can be found at
  http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
  was generated from part of bacterial clone contigs of human
  chromosome 10, constructed by the Sanger Centre Chromosome 10
  Mapping Group. Further information can be found at
  http://www.sanger.ac.uk/HGP/Chr10
  Rpl1-383C6 is from the library RPEC1-11.2 constructed by the group
  of Plier de Jong. For further details see
  http://www.chori.org/bacpac/home.htm
  VECTOR: pBACe3.6
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: vegas@sanger.ac.uk
  -----
  This sequence was finished as follows unless otherwise noted: all
  regions were either double-stranded or sequenced with an alternate
  chemistry or covered by high quality data (i.e., phred quality >=
  30); an attempt was made to resolve all sequencing problems, such
  as compressions and repeats; all regions were covered by at least
  one subclone; and the assembly was confirmed by restriction digest,
  except on the rare occasion of the clone being a YAC.

FEATURES
  source
  1..112442
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="RZPD:RPCIB753C06383"
  /db_xref="taxon:9606"
  /chromosome="10"
  /clone="RPl1-383C6"
  /clone_1fb="RPEC1-11.2"
  /clone_2fb="RPEC1-11.2"
  misc_feature
  /note="Clome right end: Rpl1-106M7"
  join(24099..24445,28551..28649,37000..37055,45632..45713,
  56274..56346,58041..58180,60225..60291,61029..63577)
  /gene="TRUB1"
  /locus_tag="Rpl1-383C6.1-001"

```



|||||  
Db 88810 ATTATGTTTCTCAGATCTTTTATTTTATTTTGAACAGAGTCTCATTGTCA 88869  
1260 CCCAGGCTGAGTACAGTGGGCTGTGCTCCGGCTCAGTGCACCTCTGCTCCAGTTTC 1319  
Db 88870 CCCAGGCTGAGTACAGTGGGCTGTGCTCCGGCTCAGTGCACCTCTGCTCCAGTTTC 88929  
Qy 1320 AAGGATCTCTCCTGACGCTTCCGAAATAGCTGGATTACAGGCGGACCAACATG 1379  
Db 88930 AAGGATCTCTCCTGACGCTTCCGAAATAGCTGGATTACAGGCGGACCAACATG 88989  
Qy 1380 CTAATTTTGTATTTTATTTAGTACAGACAGATTTCCTCATTTGACAGGCTTGCCTTGA 1439  
Db 88990 CTAATTTTGTATTTTATTTAGTACAGACAGATTTCCTCATTTGACAGGCTTGCCTTGA 89049  
Qy 1440 CTCCTGACTTACAGGATCCACCCACCTCAGCTCCCAAGACCTGGGATTACAGGATC 1499  
Db 89050 CTCCTGACTTACAGGATCCACCCACCTCAGCTCCCAAGACCTGGGATTACAGGATC 89109  
Qy 1500 AGCCACCGTCCGACCTGTCTTCTCAGATCTCTGA- TTTGTTCTGAAGCCTTCATTTC 1558  
Db 89110 AGCCACCGTCCGACCTGTCTTCTCAGATCTCTGA TTTGTTCTGAAGCCTTCATTTC 89169  
Qy 1559 TATCTTCTTATTTTATTTTGAAGTGTACCTTAAGTAAAGTTTAAACAATCAATATC 1618  
Db 89170 TATCTTCTTATTTTATTTTGAAGTGTACCTTAAGTAAAGTTTAAACAATCAATATC 89229  
Qy 1619 TTTGGAAAATCCCGTCTCTCTTCTTATTTCTACAAAATATGTCAGTATAGCTGATG 1678  
Db 89230 TTTGGAAAATCCCGTCTCTCTTCTTATTTCTTAAATAATATGTCAGTATAGCTGATG 89289  
Qy 1679 TTAGTTTCTTCAATATTTATTTCTCTATCTCAGATTTATCTCATGCTCAATATGTT 1738  
Db 89290 TTAGTTTCTTCAATATTTATTTCTCTATCTCAGATTTATCTCATGCTCAATATGTT 89349  
Qy 1739 ATTGAATAGTCTTCACTTCTTGTATCTCAGTTTGTGCTCTTATTTTCACTTAAGTCTA 1798  
Db 89350 ATTGAATAGTCTTCACTTCTTGTATCTCAGTTTGTGCTCTTATTTTCACTTAAGTCTA 89409  
Qy 1799 AGTGCTATTGAATTAAGAGGCTGTGAAGATTTCTTCCAATATGTCATCTTATCTTTTG 1858  
Db 89410 AGTGCTATTGAATTAAGAGGCTGTGAAGATTTCTTCCAATATGTCATCTTATCTTTTG 89469  
Qy 1859 ACTGATGCGAGTGCACAACTGTATTAAGTGTGATTTCTTCAATATTCACAGAAACAT 1918  
Db 89470 ACTGATGCGAGTGCACAACTGTATTAAGTGTGATTTCTTCAATATTCACAGAAACAT 89529  
Qy 1919 GCTGACTCTCTCTTCTGAAAGCAATGCCCAAGCAGACATTTGATAGTATGTACG 1978  
Db 89530 GCTGACTCTCTCTTCTGAAAGCAATGCCCAAGCAGACATTTGATAGTATGTACG 89589  
Qy 1979 CAACAGGAGCATGGTGCATAGC 2001  
Db 89590 CAACAGGAGCATGGTGCATAGC 89612

RESULT 8  
AL391648/c 176432 bp DNA linear HTG 18-JUL-2001  
LOCUS AL391648 Home sapiens chromosome 10 clone RP11-203A7, 9 unordered pieces.  
DEFINITION AL391648  
ACCESSION AL391648.12 GI:1469309  
VERSION HTG: HTGS PHASE1; HTGS\_CANCELLED.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE  
1 Chapman, J.  
AUTHORS Direct Submission  
TITLE Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgehire,  
JOURNAL CB10 15A, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

requests: clonerequests@sanger.ac.uk  
On Jul 10, 2001 this sequence version replaced gi:14596357.  
-----  
Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
-----  
Project Information  
Center project name: ba203A7  
-----  
Summary Statistics  
Sequencing program: XGAP4; version 4.5  
Sequencing vector: plasmid; 108752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 174005 bases at least Q40  
Consensus quality: 174806 bases at least Q30  
Consensus quality: 175273 bases at least Q20  
Insert size: 175632; sum-of-contigs  
Insert size: 155595; 14.1% error; agarose-fp  
Quality coverage: 6.85x in Q20 bases; sum-of-contigs Quality  
Coverage: 8.13x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 23099: contig of 23099 bp in length  
2 23100: gap of 100 bp  
3 23101: contig of 14850 bp in length  
4 38049: contig of 14850 bp in length  
5 38050: gap of 100 bp  
6 38150: contig of 27281 bp in length  
7 65431: gap of 100 bp  
8 65531: contig of 35076 bp in length  
9 100607: gap of 100 bp  
10 100706: gap of 100 bp  
11 156710: contig of 56004 bp in length  
12 156711: gap of 100 bp  
13 156811: contig of 2423 bp in length  
14 159233: contig of 2423 bp in length  
15 159234: gap of 100 bp  
16 159334: contig of 12399 bp in length  
17 171733: gap of 100 bp  
18 171832: gap of 100 bp  
19 171833: gap of 2263 bp in length  
20 174095: contig of 2263 bp in length  
21 174096: gap of 100 bp  
22 174196: contig of 2237 bp in length.  
-----  
FEATURES  
source  
1..176432  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-203A7"  
/clone\_lib="RPCT-11.1"  
1..23099  
/note="assembly\_fragment:01553  
fragment\_chain:1  
clone\_end:77  
vector\_side:left"  
23200..38049  
/note="assembly\_fragment:00834  
fragment\_chain:1"  
38150..65430  
/note="assembly\_fragment:02112  
fragment\_chain:1"  
65531..100606  
/note="assembly\_fragment:01440  
fragment\_chain:1"  
100707..156710  
/note="assembly\_fragment:01209  
fragment\_chain:1"  
156811..159233  
/note="assembly\_fragment:00074"

misc\_feature 159334..171732  
/note="assembly\_fragment:01455"  
misc\_feature 171833..174095  
/note="assembly\_fragment:01550"  
misc\_feature 174196..176432  
/note="assembly\_fragment:03092"

ORIGIN

Query Match 98.7%; Score 1974.2; DB 14; Length 176432;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1998; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 AATGTTGGATTTCAAAGGTAACAAGAACTGGAAAAACAACACTAGCGCGAGTCTAGAG 60  
Db 26028 AATGTTGGATTTCAAAGGTAACAAGAACTGGAAAAACAACACTAGCGCGAGTCTAGAG 25965  
QY 61 TTTACGCGAGACTGCTGACGCTTGTGTTTTTCACTGACAGCTGAAAATGAGCCACG 120  
Db 25968 TTTACGCGAGACTGCTGACGCTTGTGTTTTTCACTGACAGCTGAAAATGAGCCACG 25909  
QY 121 TTCAGGAAAGCTTTGCTTCCCTCCCTCAAGGTTACCAAACTCTCAGTCTCTCAGG 180  
Db 25908 TTCAGGAAAGCTTTGCTTCCCTCCCTCAAGGTTACCAAACTCTCAGTCTCTCAGG 25849  
QY 181 AAAGCCAAAAATGAATTTAGAGGTTAGAGTTGTTCTTTATCTATTAACAGATTG 240  
Db 25848 AAAGCCAAAAATGAATTTAGAGGTTAGAGTTGTTCTTTATCTATTAACAGATTG 25789  
QY 241 AATAATATGTTCTCTCAACAGATGTTCTGTTTAAACAATACTCACTTCTGACACTAG 300  
Db 25788 AATAATATGTTCTCTCAACAGATGTTCTGTTTAAACAATACTCACTTCTGACACTAG 25729  
QY 301 CATATGCGAGAGTGTACTACCAAGGTAAACAAGATGCGTCCCAATTCGAATCCC 360  
Db 25728 CATATGCGAGAGTGTACTACCAAGGTAAACAAGATGCGTCCCAATTCGAATCCC 25669  
QY 361 TGAATGAGTGAAGAAATCAGAAATTATATAGGGGATTCACAGAGCTGGCTACGAGT 420  
Db 25668 TGAATGAGTGAAGAAATCAGAAATTATATAGGGGATTCACAGAGCTGGCTACGAGT 25609  
QY 421 TGCAGTGGTCAGATACTTGTCTCATATACGACAGGTGCTGCTGTATGACAACTGCTCA 480  
Db 25608 TGCAGTGGTCAGATACTTGTCTCATATACGACAGGTGCTGCTGTATGACAACTGCTCA 25549  
QY 481 CTGCTTCATTTCTGCTGCTGCTTCTTAATAACTGCTTTTCTCAGCTCAATGGCTTTCTT 540  
Db 25548 CTGCTTCATTTCTGCTGCTGCTTCTTAATAACTGCTTTTCTCAGCTCAATGGCTTTCTT 25489  
QY 541 CCTCTGGCAGTCAAGTTTCTTTGGGTCAAACAGCAAAATGATCTTTAGATCACCTGGT 600  
Db 25488 CCTCTGGCAGTCAAGTTTCTTTGGGTCAAACAGCAAAATGATCTTTAGATCACCTGGT 25429  
QY 601 ACTCAAGAGCTACAGACATTTGGGCACTTCACCTCTCTTGGAAAAACAATTTTA 660  
Db 25428 ACTCAAGAGCTACAGACATTTGGGCACTTCACCTCTCTTGGAAAAACAATTTTA 25369  
QY 661 TGGAAAGCCAGGTTGCTCATGCTCTTTGAGGTTGTTTCTCAGCGCAAGGCCAAGCTT 720  
Db 25368 TGGAAAGCCAGGTTGCTCATGCTCTTTGAGGTTGTTTCTCAGCGCAAGGCCAAGCTT 25309  
QY 721 TGTGCTTCAAAATGAAATTTAGAGGCTTCAGAACAGATGCACATTTTCATATGCGCTCA 780  
Db 25308 TGTGCTTCAAAATGAAATTTAGAGGCTTCAGAACAGATGCACATTTTCATATGCGCTCA 25249  
QY 781 CCCAACTGGATTAAGAAACAATTTGCTATCTCAATGACCACTTTT-TCAGGTGGAGT 839  
Db 25248 CCCAACTGGATTAAGAAACAATTTGCTATCTCAATGACCACTTTTCTCAGGTGGAGT 25189  
QY 840 GTAGATGCTGGAATGGTTCACAGCAATTCGCCAACCAAACTTTGGAAAAAGGCTGGAAGC 899  
Db 25188 GTAGATGCTGGAATGGTTCACAGCAATTCGCCAACCAAACTTTGGAAAAAGGCTGGAAGC 25129  
QY 900 TCTGACTGGGAGCCCTAATATATGCAAAAGTTGATAGGCTCTTCATGCAAAATATGAACC 959

Db 25128 TCTGACTGGGAGCCCTAATATATGCAAAAGTTATAGGCTCTTCATGCAAAATATGAACC 25069  
QY 960 CGTGTATGATATAGCTAAAGGTTGGCTTTATGTTTCTATTCCTTCCACAACTGGTA 1019  
Db 25068 CGTGTATGATATAGCTAAAGGTTGGCTTTATGTTTCTATTCCTTCCACAACTGGTA 25009  
QY 1020 GAATGATATGCTTGTGTTCCCTTTAAAAATGCAACAATGCAATTTATAGTCTGTGA 1079  
Db 25008 GAATGATATGCTTGTGTTCCCTTTAAAAATGCAACAATTTATAGTCTGTGA 24949  
QY 1080 TAGTAACACAGATCATGCTCCATGAATAATGCTTCAGAACCCAAATATAGAGATTTT 1139  
Db 24948 TAGTAACACAGATCATGCTCCATGAATAATGCTTCAGAACCCAAATATAGAGATTTT 24889  
QY 1140 TAGCAATGTGACAAAGAGGCCATTTCACTGTGTAATTTGTTCAAGAAATTTTG 1199  
Db 24888 TAGCAATGTGACAAAGAGGCCATTTCACTGTGTAATTTGTTCAAGAAATTTTG 24829  
QY 1200 ATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCACTTGTGA 1259  
Db 24828 ATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCACTTGTGA 24769  
QY 1260 CCCAGCTGAGATCAGTGGCTGTGCTCGGCTCACTGCAACTCTGCTCCAGGTTT 1319  
Db 24768 CCCAGCTGAGATCAGTGGCTGTGCTCGGCTCACTGCAACTCTGCTCCAGGTTT 24709  
QY 1320 AAGCATTTCTCTGACGCTTCCCAATATGCTGGAGTTACAGGCGCATCACCACTGC 1379  
Db 24708 AAGCATTTCTCTGACGCTTCCCAATATGCTGGAGTTACAGGCGCATCACCACTGC 24649  
QY 1380 CTAAATTTTGTATTTTATAGTAAAGACAGATTTGCGCATTTTGACAGGCTTGCTTGA 1439  
Db 24648 CTAAATTTTGTATTTTATAGTAAAGACAGATTTGCGCATTTTGACAGGCTTGCTTGA 24589  
QY 1440 CTCTGACTTCAGTGAATCCACCACCTCAGGCTCCCAAGGCTGGATTCAGGCAATG 1499  
Db 24588 CTCTGACTTCAGTGAATCCACCACCTCAGGCTCCCAAGGCTGGATTCAGGCAATG 24529  
QY 1500 AGCCACCGTGCCAGCTGTGTTTCTCAGATCTGTGA-TTTTGTTTGAAGCTTCATTTT 1558  
Db 24528 AGCCACCGTGCCAGCTGTGTTTCTCAGATCTGTGAATTTGTTTGAAGCTTCATTTT 24469  
QY 1559 TATCTTTATATCATTTTGAAGTGTACACCTAGTAAAGTTTAAACAATCAATATTC 1618  
Db 24468 TATCTTTATATCATTTTGAAGTGTACACCTAGTAAAGTTTAAACAATCAATATTC 24409  
QY 1619 TTTGAAAAATTCCTGTTCTCTTTCTATATCTCAAAAAATATGTCAGTATAGCTGATG 1678  
Db 24408 TTTGAAAAATTCCTGTTCTCTTTCTATATCTCAAAAAATATGTCAGTATAGCTGATG 24349  
QY 1679 TTATGTTTCTTCAAAATATTCATTTCTATCTCAGAAATTTATCTCATATGTT 1738  
Db 24348 TTATGTTTCTTCAAAATATTCATTTCTATCTCAGAAATTTATCTCATATGTT 24289  
QY 1739 ATTGAATAGTCTCACTTCTGTATCCAGTTTCTGTCTCTTATTTCACTCTAAGTCTTA 1798  
Db 24288 ATTGAATAGTCTCACTTCTGTATCCAGTTTCTGTCTCTTATTTCACTCTAAGTCTTA 24229  
QY 1799 AGTGCCTATTGAATTAAGAGCTTGTAAACAGATCTTTCTCCAAATATGCTTATCTTTTG 1858  
Db 24228 AGTGCCTATTGAATTAAGAGCTTGTAAACAGATCTTTCTCCAAATATGCTTATCTTTTG 24169  
QY 1859 ACTGCATCCAGTGAACAATGTTAATCTGTTTGAATCTTCAATTCACAGAAAT 1918  
Db 24168 ACTGCATCCAGTGAACAATGTTAATCTGTTTGAATCTTCAATTCACAGAAAT 24109  
QY 1919 GCTGATCTCTCTCTTCCGAAAGCAATGCCAAGACAGCAAGTGTATATAGTATAGTACG 1978  
Db 24108 GCTGATCTCTCTCTTCCGAAAGCAATGCCAAGACAGCAAGTGTATATAGTATAGTACG 24049  
QY 1979 CAACAGGACATGGGTGATAGC 2001

Db 24048 CAACAGGACATGGGTGCATAGC 24026

RESULT 9  
AC084084  
LOCUS

DEFINITION  
AC084084 71032 bp DNA linear HTG 25-MAR-2001  
Homo sapiens chromosome 8 clone RP11-314C19 map 8, LOM-PASS  
SEQUENCE SAMPLING.

AC084084  
AC084084.2 GI:13446278  
HTG: HTGS PHASE0  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 71032)  
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 8, clone RP11-314C19  
Unpublished  
2 (bases 1 to 71032)  
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Adirham,H., Allen,N.,  
Anderson,S., Barra,N., Bastien,V., Beda,F., Boguslavskiy,L.,  
Bouhagiel,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearrellan,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Labrecque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPhaetere,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., O'Neill,D., Olivart,T.M., Oliver,J., Peterson,K.,  
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirelli,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
Zimmer,A. and Zody,M.

Direct Submission  
Submitted (12-Oct-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 25, 2001 this sequence version replaced gi:10799449.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L11327  
Center clone name: 314\_C19

-----  
\* NOTE: This record contains 85 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 720: contig of 720 bp in length  
\* 721 820: gap of 100 bp  
\* 821 1557: contig of 737 bp in length  
\* 1558 1657: gap of 100 bp  
\* 1658 2404: contig of 747 bp in length

\* 2405 2504: gap of 100 bp  
\* 2505 3299: contig of 795 bp in length  
\* 3300 3399: gap of 100 bp  
\* 3400 4153: contig of 754 bp in length  
\* 4154 4253: gap of 100 bp  
\* 4254 4978: contig of 725 bp in length  
\* 4979 5078: gap of 100 bp  
\* 5079 5815: contig of 737 bp in length  
\* 5816 5915: gap of 100 bp  
\* 5916 6648: contig of 733 bp in length  
\* 6649 6748: gap of 100 bp  
\* 6749 7520: contig of 772 bp in length  
\* 7521 7620: gap of 100 bp  
\* 7621 8347: contig of 727 bp in length  
\* 8348 8447: gap of 100 bp  
\* 8448 9227: contig of 780 bp in length  
\* 9228 9327: gap of 100 bp  
\* 9328 10052: contig of 725 bp in length  
\* 10053 10152: gap of 100 bp  
\* 10153 10869: contig of 717 bp in length  
\* 10870 10969: gap of 100 bp  
\* 10970 11718: contig of 749 bp in length  
\* 11719 11818: gap of 100 bp  
\* 11819 12557: contig of 739 bp in length  
\* 12558 12657: gap of 100 bp  
\* 12658 14412: contig of 755 bp in length  
\* 13413 13512: gap of 100 bp  
\* 13513 14255: contig of 743 bp in length  
\* 14256 14355: gap of 100 bp  
\* 14356 15086: contig of 731 bp in length  
\* 15087 15187: gap of 100 bp  
\* 15188 15901: contig of 715 bp in length  
\* 15902 16001: gap of 100 bp  
\* 16002 16711: contig of 710 bp in length  
\* 16712 16811: gap of 100 bp  
\* 16813 17539: contig of 728 bp in length  
\* 17540 17639: gap of 100 bp  
\* 17640 18351: contig of 712 bp in length  
\* 18352 18451: gap of 100 bp  
\* 18452 19162: contig of 711 bp in length  
\* 19163 19262: gap of 100 bp  
\* 19263 20054: contig of 792 bp in length  
\* 20055 20154: gap of 100 bp  
\* 20155 20891: contig of 737 bp in length  
\* 20892 20991: gap of 100 bp  
\* 20992 21720: contig of 729 bp in length  
\* 21721 21820: gap of 100 bp  
\* 21821 22553: contig of 733 bp in length  
\* 22554 22653: gap of 100 bp  
\* 22654 23420: contig of 767 bp in length  
\* 23421 23520: gap of 100 bp  
\* 23521 24234: contig of 714 bp in length  
\* 24235 24334: gap of 100 bp  
\* 24335 25069: contig of 735 bp in length  
\* 25070 25169: gap of 100 bp  
\* 25170 25915: contig of 746 bp in length  
\* 25916 26015: gap of 100 bp  
\* 26016 26751: contig of 736 bp in length  
\* 26752 26851: gap of 100 bp  
\* 26852 27574: contig of 723 bp in length  
\* 27575 27674: gap of 100 bp  
\* 27675 28437: contig of 763 bp in length  
\* 28438 28537: gap of 100 bp  
\* 28538 29258: contig of 721 bp in length  
\* 29259 29358: gap of 100 bp  
\* 29359 30068: contig of 710 bp in length  
\* 30069 30168: gap of 100 bp  
\* 30169 30918: contig of 750 bp in length  
\* 30919 31018: gap of 100 bp  
\* 31019 31770: contig of 752 bp in length  
\* 31771 31870: gap of 100 bp  
\* 31871 32658: contig of 788 bp in length  
\* 32659 32758: gap of 100 bp

```
* 32759 33501: contig of 743 bp in length
* 33502 33601: gap of 100 bp
* 33602 34347: contig of 746 bp in length
* 34348 34447: gap of 100 bp
* 34448 35165: contig of 718 bp in length
* 35166 35265: gap of 100 bp
* 35266 35998: contig of 733 bp in length
* 35999 36098: gap of 100 bp
* 36099 36802: contig of 704 bp in length
* 36803 36902: gap of 100 bp
* 36903 37632: contig of 730 bp in length
* 37633 37732: gap of 100 bp
* 37733 38454: contig of 722 bp in length
* 38455 39275: gap of 100 bp
* 39276 39375: gap of 100 bp
* 39376 40113: contig of 738 bp in length
* 40114 40213: gap of 100 bp
* 40214 40948: contig of 735 bp in length
* 40949 41048: gap of 100 bp
* 41049 41892: contig of 744 bp in length
* 41893 41992: gap of 100 bp
* 41993 42630: contig of 738 bp in length
* 42631 42730: gap of 100 bp
* 42731 43459: contig of 729 bp in length
* 43460 43559: gap of 100 bp
* 43560 44277: contig of 718 bp in length
* 44278 44377: gap of 100 bp
* 44378 45120: contig of 743 bp in length
* 45121 45220: gap of 100 bp
* 45221 45942: contig of 722 bp in length
* 45943 46042: gap of 100 bp
* 46043 46775: contig of 733 bp in length
* 46776 46875: gap of 100 bp
* 46876 47709: contig of 734 bp in length
* 47710 48446: gap of 100 bp
* 48447 48546: gap of 100 bp
* 48547 49285: contig of 739 bp in length
* 49286 49385: gap of 100 bp
* 49386 50178: contig of 793 bp in length
* 50179 50279: gap of 100 bp
* 50279 51008: contig of 730 bp in length
* 51009 51108: gap of 100 bp
* 51109 51846: contig of 738 bp in length
* 51847 51946: gap of 100 bp
* 51947 52672: contig of 726 bp in length
* 52673 52772: gap of 100 bp
* 52773 53505: contig of 733 bp in length
* 53506 53605: gap of 100 bp
* 53606 54330: contig of 725 bp in length
* 54331 54430: gap of 100 bp
* 54431 55166: contig of 736 bp in length
* 55167 55266: gap of 100 bp
* 55267 55979: contig of 713 bp in length
* 55980 56079: gap of 100 bp
* 56080 56823: contig of 744 bp in length
* 56824 56923: gap of 100 bp
* 56924 57626: contig of 703 bp in length
```

Query Match 33.5%; Score 670.4; DB 14; Length 71032;

Best Local Similarity 87.0%; Pred. No. 5.7e-142;

Matches 693; Conservative 0; Mismatches 102; Indels 2; Gaps 2;

```
QY 131 CTGTGTTCTTCCCTCCCTCAAGTTACCCACATTTCTCAAGTTCTCAGAAAAGCCAAA 190
Db 41937 CTGTGTTCTTCCCTCCCTC-AGGTTAACCCACATTTCTCAAGTTCTCAGAAAAGCCAAA 41995
QY 191 AATGAATTTGAGGTTTAGGATTGTGGTTCTTTATATATATACAGATTGATTAATATGTT 250
Db 41996 AATGAATTTGAGGTTTAGGATTGTGGTTCTTTATATATATACAGATTGATTAATATGTT 42055
QY 251 CCTCCACACAGATGTTGTGTTGTAACAATACTCACTTCTGACACATGCAATATGACAG 310
```

```
Db 42056 CCTCCACACAGATGTTGTGTTGTAACAATACTCACTTCTGACACATGCAATATGACAG 42115
QY 311 AGTGTACTACCAAGATAACACAGATTGGCTGCCCAATTCCTGAACATGAGT 370
Db 42116 AGTGTACTACCAAGATAACACAGATTGGCTGCCCAATTCCTGAACATGAGT 42175
QY 371 GAGAGAAATCAGAAATTAATATAGGGGATTCACACAGAGCTGGCTAGAGATGTCAGTGT 430
Db 42176 GAGAGAAATCAGAAATTAATATAGGGGATTCACACAGAGCTGGCTAGAGATGTCAGTGT 42235
QY 431 CAGATACCTTGGCTATCATACGACAGTGTGCTGTAGCAACTGCTCACTGCTTCATT 490
Db 42236 CAGATACCTTGGCTATCATACGACAGTGTGCTGTAGCAACTGCTCACTGCTTCATT 42295
QY 491 TCTGCTGTTGGCTTTAAATATCTGCTTTTCTAGCTCAATGGCTTTCTCCCTGCGCA 550
Db 42296 TCTGCTGTTGGCTTTAAATATCTGCTTTTCTAGCTCAATGGCTTTCTCCCTGCGCA 42355
QY 551 GTACAGTTCTTGGGTCACACAGAAATGATTTCTTTGAATCAGCTGCTACCAAGGA 610
Db 42356 GTACAGTTCTTGGGTCACACAGAAATGATTTCTTTGAATCAGCTGCTACCAAGGA 42415
QY 611 GCTACAGACATTTGGGATCCCACTTCCACTCTCTTGGAAAAACAATTTATGGAAGCCAA 670
Db 42416 GCTACAGACATTTGGGATCCCACTTCCACTCTCTTGGAAAAACAATTTATGGAAGCCAA 42475
QY 671 GGTTGCGATAGTGCTCTT-GAGGTGTTTGTCTACGCCAAGGCCAAGCTTGTGCTTCA 729
Db 42476 GGTTGCGATAGTGCTCTTNGAGGTGTTGTCTACGCCAAGGCCAAGCTTGTGCTTCA 42535
QY 730 AACATGAATTTGAGAGCTTCAGAACAGATTCATTTGCAATGCGCTCACCAACTG 789
Db 42536 AACATGAATTTGAGAGCTTCAGAACAGATTCATTTGCAATGCGCTCACCAACTG 42595
QY 790 ATAAAAGAACATTTGCGATATCTCAATGACCACTTTTTCAGGTGAGATGATAGTCTG 849
Db 42596 ATAAAAGAACATTTGCGATATCTCAATGACCACTTTTTCAGGTGAGATGATAGTCTG 42655
QY 850 GAATGGTCACAGATTCGCCAACCAAACTTTGCAAAAAAGGCTGGAAGCTGACTGAG 909
Db 42656 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 42715
QY 910 GACCTTAATATGCAA 926
Db 42716 NNNNNNNNNNNNNNNNGA 42732
```

RESULT 10

AP006290/c 27423 bp DNA linear PRI 04-APR-2003

LOCUS Homo sapiens genomic DNA, chromosome 8 clone:RP11-1345M15, complete sequence.

AP006290

AP006290.2 GI:29539558

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y. Home sapiens genomic DNA Published Only in Database (2003)

2 (bases 1 to 27423)

Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y. Direct Submission Submitted (24-MAR-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan







1	654: contig of 654 bp in length
655	754: gap of 100 bp
755	1683: contig of 929 bp in length
1684	1783: gap of 100 bp
1784	2094: contig of 311 bp in length
2095	2194: gap of 100 bp
2195	3125: contig of 931 bp in length
3126	3225: gap of 100 bp
3226	4390: contig of 1065 bp in length
4391	4390: gap of 100 bp
4391	5051: contig of 661 bp in length
5052	5151: gap of 100 bp
5152	7255: contig of 2105 bp in length
7255	7356: gap of 100 bp
7357	8674: contig of 1318 bp in length
8675	8774: gap of 100 bp
8775	10666: contig of 1892 bp in length
10667	10766: gap of 100 bp
10767	12819: contig of 2053 bp in length
12820	12919: gap of 100 bp
12920	14779: contig of 1860 bp in length
14780	14879: gap of 100 bp
14880	17960: contig of 3081 bp in length
17961	18060: gap of 100 bp
18061	21424: contig of 3364 bp in length
21425	21524: gap of 100 bp
21525	23085: contig of 1561 bp in length
23086	23185: gap of 100 bp
23186	28363: contig of 5178 bp in length
28364	28463: gap of 100 bp
28464	31735: contig of 3272 bp in length
31736	31835: gap of 100 bp
31836	39013: contig of 7178 bp in length
39014	39113: gap of 100 bp
39114	45836: contig of 6723 bp in length
45837	45936: gap of 100 bp
45937	65312: contig of 17376 bp in length
65313	65312: gap of 100 bp
65313	71785: contig of 8373 bp in length
71786	71885: gap of 100 bp
71886	82043: contig of 10158 bp in length
82044	82143: gap of 100 bp
82144	92091: contig of 9948 bp in length
92092	92191: gap of 100 bp
92192	106557: contig of 14366 bp in length
106558	106557: gap of 100 bp
106658	133333: contig of 26656 bp in length.

URES	Location/Qualifiers
source	1. .133315
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="8"
	/clone="RP11-933110"
misc_feature	1. .654
	/note="assembly_fragment"
gap	655. .754
	/estimated_length=100
misc_feature	755. .1683
	/note="assembly_fragment"
gap	1684. .1783
	/estimated_length=100
misc_feature	1784. .2094
	/note="assembly_fragment"
gap	2095. .2194
	/estimated_length=100
misc_feature	2195. .3125
	/note="assembly_fragment"
gap	3126. .3225
	/estimated_length=100
misc_feature	3226. .4290

gap	/note="assembly_fragment"
	4231. .4390
misc_feature	/estimated_length=100
	4391. .5051
gap	/note="assembly_fragment"
	5052. .5151
misc_feature	/estimated_length=100
	5152. .7256
gap	/note="assembly_fragment"
	7257. .7356
misc_feature	/estimated_length=100
	7357. .8674
gap	/note="assembly_fragment"
	8675. .8774
misc_feature	/estimated_length=100
	8775. .10666
gap	/note="assembly_fragment"
	10667. .10766
misc_feature	/estimated_length=100
	10767. .12819
gap	/note="assembly_fragment"
	12820. .12919
misc_feature	/estimated_length=100
	12920. .14779
misc_feature	/note="assembly_fragment"
gap	14780. .14879
	/estimated_length=100
misc_feature	14880. .17960
gap	/note="assembly_fragment"
	17961. .18060
misc_feature	/estimated_length=100
	18061. .21424
gap	/note="assembly_fragment"
	21425. .21524
misc_feature	/estimated_length=100
	21525. .22085

Query Match 10.98; Score 218; DB 14; Length 133315;

Best Local Similarity 68.4%; Pred. No. 3.6e-39;  
Matches 361; Conservative 0; Mismatches 150. Indels 7; Gaps 4.

DATE	DESCRIPTION	AMOUNT	CHECK NO.	BANK
1/1/58	DEPOSIT	100.00		CHASE
1/15/58	PAYROLL	50.00	101	CHASE
1/30/58	RENT	25.00	102	CHASE
2/1/58	DEPOSIT	100.00		CHASE
2/15/58	PAYROLL	50.00	103	CHASE
2/28/58	RENT	25.00	104	CHASE
3/1/58	DEPOSIT	100.00		CHASE
3/15/58	PAYROLL	50.00	105	CHASE
3/30/58	RENT	25.00	106	CHASE
4/1/58	DEPOSIT	100.00		CHASE
4/15/58	PAYROLL	50.00	107	CHASE
4/30/58	RENT	25.00	108	CHASE
5/1/58	DEPOSIT	100.00		CHASE
5/15/58	PAYROLL	50.00	109	CHASE
5/30/58	RENT	25.00	110	CHASE
6/1/58	DEPOSIT	100.00		CHASE
6/15/58	PAYROLL	50.00	111	CHASE
6/30/58	RENT	25.00	112	CHASE
7/1/58	DEPOSIT	100.00		CHASE
7/15/58	PAYROLL	50.00	113	CHASE
7/30/58	RENT	25.00	114	CHASE
8/1/58	DEPOSIT	100.00		CHASE
8/15/58	PAYROLL	50.00	115	CHASE
8/30/58	RENT	25.00	116	CHASE
9/1/58	DEPOSIT	100.00		CHASE
9/15/58	PAYROLL	50.00	117	CHASE
9/30/58	RENT	25.00	118	CHASE
10/1/58	DEPOSIT	100.00		CHASE
10/15/58	PAYROLL	50.00	119	CHASE
10/30/58	RENT	25.00	120	CHASE
11/1/58	DEPOSIT	100.00		CHASE
11/15/58	PAYROLL	50.00	121	CHASE
11/30/58	RENT	25.00	122	CHASE
12/1/58	DEPOSIT	100.00		CHASE
12/15/58	PAYROLL	50.00	123	CHASE
12/30/58	RENT	25.00	124	CHASE
1/1/59	DEPOSIT	100.00		CHASE
1/15/59	PAYROLL	50.00	125	CHASE
1/30/59	RENT	25.00	126	CHASE
2/1/59	DEPOSIT	100.00		CHASE
2/15/59	PAYROLL	50.00	127	CHASE
2/28/59	RENT	25.00	128	CHASE
3/1/59	DEPOSIT	100.00		CHASE
3/15/59	PAYROLL	50.00	129	CHASE
3/30/59	RENT	25.00	130	CHASE
4/1/59	DEPOSIT	100.00		CHASE
4/15/59	PAYROLL	50.00	131	CHASE
4/30/59	RENT	25.00	132	CHASE
5/1/59	DEPOSIT	100.00		CHASE
5/15/59	PAYROLL	50.00	133	CHASE
5/30/59	RENT	25.00	134	CHASE
6/1/59	DEPOSIT	100.00		CHASE
6/15/59	PAYROLL	50.00	135	CHASE
6/30/59	RENT	25.00	136	CHASE
7/1/59	DEPOSIT	100.00		CHASE
7/15/59	PAYROLL	50.00	137	CHASE
7/30/59	RENT	25.00	138	CHASE
8/1/59	DEPOSIT	100.00		CHASE
8/15/59	PAYROLL	50.00	139	CHASE
8/30/59	RENT	25.00	140	CHASE
9/1/59	DEPOSIT	100.00		CHASE
9/15/59	PAYROLL	50.00	141	CHASE
9/30/59	RENT	25.00	142	CHASE
10/1/59	DEPOSIT	100.00		CHASE
10/15/59	PAYROLL	50.00	143	CHASE
10/30/59	RENT	25.00	144	CHASE
11/1/59	DEPOSIT	100.00		CHASE
11/15/59	PAYROLL	50.00	145	CHASE
11/30/59	RENT	25.00	146	CHASE
12/1/59	DEPOSIT	100.00		CHASE
12/15/59	PAYROLL	50.00	147	CHASE
12/30/59	RENT	25.00	148	CHASE
1/1/60	DEPOSIT	100.00		CHASE
1/15/60	PAYROLL	50.00	149	CHASE
1/30/60	RENT	25.00	150	CHASE</

1216 TCTTTTATTATTTTGTGAACAGAGTCTCACTTGTCAACCCAGGCTGAGTACA 1275

41158 TTTTCTTTTCTGACACAGATTCTACTGTGACACCCAGGCTGAGTCA 4109

[illegible]

1276 GTGGCTGTGTCTCTCGCTACCTGCAACCTCTCTCTCCAGGTTCAAGCGATTCTCTG-- 1333

41098 GTGAC-ACGATCTGGCTCACTGCACCTCTGCGCTCCAGGTTCAAGTGTCTCTCC 4104

1 3 3 4 T T C C A A T T A A C C C C A A T T C C A A C C C C A A T T C C - C A T A A T T T A T T T T C C T T A 1 3 3 1

[illegible]

41039 TCAGCCTCCGAGTAGCTGGATTACAGGCACCTGCCACCAATCCAGCTAATTTTGTA 4098

1392 TTTTAGTAGACAGAGTTTCGCCATGTTGACCAGGCTTGCCCTGAACTCCTGACTTCA 1451

409/9 TTTAGTAGAGACAGGTTCCACCAATGTTGGCAGGCTGGTCTTGAATCTTGATCTCA 4092

1452 GGTGATCCACCACCTCAGCCTCCCAAAGCACTGGGATTACAGGCATGAGCCACCGTGCC 1511

40919 AGTGATCCACCTGCTCAGCCTCCCAAAGTGTGGATTACAGCATGAGCCACTGAGCC 4086

1512 CAGCCTGTTTCTCAGATCCTGTATTGTTTCTGAAGCCTTCATTTCTATCTTCTTATTC 1571

40859 CGCCTTCCGGGTTCTTGATGATATGTTTATAATGTCAGCCTCATACAATAATAG 40860

1577 1677 1777 1877 1977 2077 2177 2277 2377 2477 2577 2677 2777 2877 2977 3077 3177 3277 3377 3477 3577 3677 3777 3877 3977 4077 4177 4277 4377 4477 4577 4677 4777 4877 4977 5077 5177 5277 5377 5477 5577 5677 5777 5877 5977 6077 6177 6277 6377 6477 6577 6677 6777 6877 6977 7077 7177 7277 7377 7477 7577 7677 7777 7877 7977 8077 8177 8277 8377 8477 8577 8677 8777 8877 8977 9077 9177 9277 9377 9477 9577 9677 9777 9877 9977

[illegible]

b 40799 TTTTACAGCTTTTACTATTTAATAAGCATTTTTCCATCATTAACATTTTAGCAAAA--C 4074

1632 CTGGTTCCTTCTTATTCCTACAAATATGTTCACTATAGCTGATGTTATGTTCTTTC 1691

4000 3000 2000 1000 0

40/41 TGGCTTTCTAATGGATGCTAAATCTCATTAATAATTAACTATCTCTCATTCGCTGGA 4068

Oy	1692	AAATGATCATTCCTCATTACCGAAGTATTCATGCCTAATGTGA	1739
Dn	40681	TATTATGTTGGTTTCAATTGTGTACATAATTAATACAATTTTATCA	40634
RESULT 12			
LOCUS	ACI30304/c		
DEFINITION	Homo sapiens chromosome 8, clone RP11-205M5, complete sequence.		
ACCESSION	ACI30304		
VERSION	ACI30304.12	GI:28467201	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	Britten,B., Nusbaum,C. and Lander,E. 1 (bases 1 to 161953) Homo sapiens chromosome 8, clone RP11-205M5 Unpublished 2 (bases 1 to 161953)		
AUTHORS	Britten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Charazo,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C., McCarthy,M., Meltrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Raymond,C., Retta,R., Riese,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnappack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamae,J., Teefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		
JOURNAL	Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
TITLE	Direct Submission		
REFERENCE	Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
AUTHORS	3 (bases 1 to 161953) Britten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Charazo,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafte,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C., McCarthy,M., Meltrim,J., Meneus,L., Mihova,T., Mlenga,V., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Peterson,K., Phunhkhang,P., Pierre,N., Raymond,C., Retta,R., Riese,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnappack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamae,J., Teefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		
TITLE	Submitted (03-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
JOURNAL	Submitted (03-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	4 (bases 1 to 161953) Britten,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corm-B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hafez,N., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Peterson,K., Phunhkhang,P., Pierre,N., Raymond,C., Retta,R., Riese,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnappack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamae,J., Teefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		

[illegible]

	repeat_region	/rpt_family="MER21B"	8255. .8781
	/rpt_family="MER77"	9010. .9367	
	repeat_region	/rpt_family="THE1B"	complement(9952. .10253)
	/rpt_family="AluSx"	complement(10313. .10508)	
	repeat_region	/rpt_family="AluSx"	complement(10545. .10856)
	/rpt_family="L1MC5"	10861. .10899	
	repeat_region	/rpt_family="AT_rich"	10900. .11039
	/rpt_family="FLAM_A"	11070. .11157	
	repeat_region	/rpt_family="L1MD3"	complement(11176. .11401)
	/rpt_family="L1MC5"	complement(11425. .11574)	
	repeat_region	/rpt_family="FRAM"	complement(11589. .13219)
	/rpt_family="SVA"	13320. .13265	
	repeat_region	/rpt_family="(GGGAGA)n"	complement(13412. .13711)
	/rpt_family="AluDo"	13761. .13896	
	repeat_region	/rpt_family="FLAM_C"	13898. .13998
	/rpt_family="GA-rich"	14000. .14047	
	repeat_region	/rpt_family="(CGAGA)n"	14067. .14192
	/rpt_family="AluDb"	14346. .14370	
	repeat_region	/rpt_family="(CAAAA)n"	complement(14758. .14968)
	/rpt_family="L2"	14972. .15220	
	repeat_region	/rpt_family="L1M4"	15221. .15505
	/rpt_family="AluSx"	15506. .16451	
	repeat_region	/rpt_family="L1M4"	16452. .16683
	/rpt_family="AluSx"	16684. .17223	
	repeat_region	/rpt_family="L1M4"	complement(17224. .17557)
	/rpt_family="MLT1A"	17558. .17664	
	repeat_region	/rpt_family="L1M4"	17665. .17989
	/rpt_family="AluDo"	17990. .18040	
	repeat_region	/rpt_family="L1M4"	18041. .18356
	/rpt_family="AluSq"	18357. .18447	
	repeat_region	/rpt_family="L1M4"	
Query Match	10.9%; Score 218; DB 8; Length 161953;		
Best Local Similarity	68.4%; Pred. No. 3,5e-39;		
Matches 361; Conservative 0; Mismatches 160; Indels 7; Gaps 4;			
OY	1216 TCTTTTATTTTATTTTATTTTGAACAAGATCTCACTTTGTCACCCAGGCTGAGTACA	1275	
Db	74986 TTTTATTTTATTTTCTTTTGTGAGACAGATTCCTCACTGTGACACCCAGGCTGAGTACA	74927	
OY	1276 GTGCTGTGATTCGGCTCACTGCAACCTTGCTGCCAGGTTCAGGCAATTCCTCTG--	1333	
Db	74926 GTGATC-ACGATCTCGGCTCACTGCAACCTTGCTGCCAGGTTCAGGCAATTCCTCTGCC	74868	

OY		133	TGAGCTCCCGAATPAGCTGGGAATTACAGGCGCATGTACACCAATGC--CTAAATTTTGTAA	1391
Dd		74867	TCAGCTCCTCCGAGTAGCTGGGAATTACAGGACTGCCACCAATCAAGCAATTTTTTGTAA	74806
OY		1392	TTTTTAAGTAAGAACAAGATTTCGCCAATGTTGACCAAGCTTGCTTGACCTCTGACTTCA	1451
Dd		74807	TTTTTATATAGAAGACAGGGTTTCCACAAATGTTGGCAGAGCTGTGTATACTCGAGCTCA	74748
OY		1452	GGTGATCACCCACCTCAGCTTCCCAGAACGCTGGGATTAACAGGCATGAGCCAACGCTGCC	1511
Dd		74747	AGTGATCCACCTGCTCAGCCTCCCAAAGCTGGGATTAACAGGCAATGAGCCACTGAGCC	74688
OY		1512	CAGCTGTTTTCTCAAGATCCGTAAATTTGTTTCTGAAGCCCTCATTTCTPATCTTCTAATTC	1571
Dd		74687	CGGCTCTCTGGGGTCTTGATGATATGTTTTATTAATGTCAGCCTCATACAAAATATATAG	74622
OY		1572	ATTTTGAGATAGTACACCTAAGTAAGTTTTTTTAAATCAATCAATATCTTTGGAAAAATCC	1631
Dd		74627	TTTTTACAGCTTTTATGATTTATTAATGAACATTTTTCATCATCAATTAACATTTTGAAGAAA--C	74570
OY		1632	CTGGTCTCTTCTTATTTCTTACAAAAATATGTTCAAGTATAGCTGATGTTATGTTCTTTTC	1691
Dd		74569	TTCGTTTTCTTAAAGGATGCTTAAAACTCATATATATTTTAACTATATCTCTATTCGTGGA	74510
OY		1692	AAATATATCATTTCTGATCTCAGAAATTAATCTCATGCCATATGTTTA	1739
Dd		74509	TATTTATGTTGTTTCAATTTGTACAAATATATATAACATTTTATCA	74462
RESULT 13				
AL353716/c				
LOCUS				
DEFINITION		AL353716	98340 bp	linear PRI 18-MAY-2005
				Human DNA sequence from clone RP11-501118 on chromosome 6 Contains
				the TBCC gene for tubulin-specific chaperone c, part of a novel
				gene, part of the gene for a novel protein (KIAA0240) and Cpg
				islands, complete sequence.
ACCESSION		AL353716		
VERSION		AL353716.18	GI:11967504	
KEYWORDS		HTG; KIAA0240; TBCC; tubulin.		
SOURCE		Homo sapiens		
ORGANISM		Homo sapiens (human)		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
		Homidae; Homo.		
		1 (bases 1 to 98340)		
REFERENCE		Laird,G.		
AUTHORS		Direct Submission		
TITLE		Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,		
JOURNAL		Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk		
		clone requests: clonerequest@sanger.ac.uk		
		On Dec 22, 2000 this sequence version replaced gi:11611339.		
		The following abbreviations are used to associate primary accession		
		numbers given in the feature table with their source databases:		
		Bm, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WormPepp; Information		
		on the WormPepp database can be found at		
		http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence		
		was generated from part of bacterial clone contigs of human		
		chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping		
		Group. Further information can be found at		
		http://www.sanger.ac.uk/HGP/Ch6		
		RP11-501118 is from the library RPCI-11.2 constructed by the group		
		of Pletier de Jong. For further details see		
		http://www.chori.org/bacpac/home.htm		
		VECTOR: pBACE3.6		
		----- Genome Center		
		Center: Wellcome Trust Sanger Institute		
		Center code: SC		
		Web site: http://www.sanger.ac.uk		
		Contact: vegas@sanger.ac.uk		
		-----		
		This sequence was finished as follows unless otherwise noted: all		
		regions were either double-stranded or sequenced with an alternate		







FEATURES	source
*	120175
*	120274: gap of 100 bp
*	120275
*	120274: contig of 8950 bp in length
*	129225
*	129324: gap of 100 bp
*	129325
*	13506: contig of 8182 bp in length
*	137507
*	137606: gap of 100 bp
*	137607
*	142607: contig of 6654 bp in length
*	144360
*	144360: gap of 100 bp
*	144361
*	150760: contig of 6400 bp in length
*	150761
*	150860: gap of 100 bp
*	150861
*	154751: contig of 3891 bp in length
*	154752
*	154851: gap of 100 bp
*	154852
*	158790: contig of 3939 bp in length
*	158791
*	158890: gap of 100 bp
*	158891
*	163714: contig of 4824 bp in length
*	163715
*	163814: gap of 100 bp
*	163815
*	166889: contig of 3175 bp in length
*	166900
*	166900: gap of 100 bp
*	169622: contig of 2533 bp in length
*	169623
*	169722: gap of 100 bp
*	169723
*	171188: contig of 1466 bp in length
*	171189
*	172289: gap of 100 bp
*	172286: contig of 1008 bp in length
	Location/Qualifiers
	1..172286

```

/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="caxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-148B3"
1..37986
/misc_feature
/note="assembly_fragment"
38087..62025
/note="assembly_fragment"
62126..73587
/note="assembly_fragment clone_end:T7 vector_side:right"
73688..84480
/note="assembly_fragment"
84581..96212
/note="assembly_fragment"
96313..105504
/note="assembly_fragment"
105605..112728
/note="assembly_fragment"
112829..120174
/note="assembly_fragment"
120275..129224
/note="assembly_fragment"
129325..137506
/note="assembly_fragment"
137607..144260
/note="assembly_fragment"
144361..150760
/note="assembly_fragment"
150861..154751
/note="assembly_fragment clone_end:SP6 vector_side:left"
154852..158790
/note="assembly_fragment"
158891..163714
/note="assembly_fragment"
163815..169889
/note="assembly_fragment"
167090..169632
/note="assembly_fragment"
169723..171188
/note="assembly_fragment"
171289..172236
/note="assembly_fragment"

```

OY		TTTGATTATGTTTTCTCAGATCTTTCATTTTATTTTATTTTGTGAACAGAATCTCACTTT	1255
Dd	73689	TTTTTTCCTTTTCTTTTCTTTTTCATTTTTTTTTTTTTTTTTTTGGACGGAGCGCTCGCTCT	73748
OY	1256	GTCACCAGGCCTGGAGNACAGTAGTGGTGTCCGGCTCACGTGCMAACTCTGCCGCCGAG	1315
Dd	73749	GTCTCCCAGGCTGGAGTGCATATGCG-ATGATCTCCGCTACATGCAACTCTGCTCTCCGG	73807
OY	1316	GTTCAACGCAATTCCTCTG--TCAGCTCCCGAATAAGCTGGAGTTACAGGCGCATGCACA	1373
Dd	73808	GTTCAACGCAATTCCTCGCTTAGCTCCCAAGTAAGCTGGAGTACAG--GCACCCACCA	73865
OY	1374	CANNCCTAA-TTTTGTGATTTTAAATNAGAACAGAGTTTCGCCATGTGACAGAGCTTTC	1433
Dd	73866	CTGTGCTGTATTTTGTATTTTAAATAGAGACGGGGTTTCCACATGTGCGCAGGCTGT	73925
OY	1434	CTTGAATCTCTGACTTCAGGTGATTCACCCACCTTCAGCTCCCAAAGCACTGGGATTACA	1493
Dd	73926	TTCGAATCTCTGATCTCAGSTATTCACCCCGCTCAGCCTCCCAAAGTCTGGGATTACA	73985
OY	1494	GGCATGAGCCACCGTGCCAGCTGTGTTTCTCAGATCTGTAATTTGTTTCT	1544
Dd	73986	GGGCTGTGACCAACGCGCCGCGCTCTTTCTTTCTTTTAAATTTT	74036

```

RESULT 17
AP000221/c
LOCUS      AP000221      89323 bp      DNA      linear      PRI 20-NOV-1999
DEFINITION Homo sapiens genomic DNA, chromosome 21q21.2, LL56-APP region,
clone:5672, complete sequence.
ACCESSION  AP000221
VERSION     AP000221
KEYWORDS   HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homidae; Homo.
REFERENCE   1 (bases 1 to 89323)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
            Homo sapiens 89,323bp genomic DNA of 21q21.2
            Published Only in DataBase (1999)
AUTHORS     2 (bases 1 to 89323)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
            Direct Submission
            Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
            Kitadato Univ., 1-15-1 Kitadato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     The sequence is a part of the data (ACCESSION No. AP000136 -
AP000145).
            The sequencing project is supported by Japan Science Technology
            Corporation (JST) and The Institute of Physical and Chemical
            Research (RIKEN).
            Location/Qualifiers
                *
                1..89323
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="21"
                /map="21q21.2"
                /clone="5672"
FEATURES
source
ORIGIN

```



QY	1167	TTTCAGGCTGGAATGTCGACGAGAACTATTGATATATGTTTTCTGAGATCTTTTAATT	1226
Db	41407	TTGAAATTGAATATTTTGTCCAGAGAAAACCTTGATTTTTTTTTTTTTTTTTTTTTTTT	41348
QY	1227	TTATTTTTTTTTGAAAACAGAGTCTCACTTTGTCAACCAGGCTGAGTACAGTGGCTGTGAT	1286
Db	41347	TTTTTTTTTTTTGAGAAAGAGTCTGCTCTGTCTCACTCAGGCTGAGTGAATGTG-TGCGAT	41288
QY	1287	CTGGGCTCACTGCAACTCTTGCCCTCCCAAGTTCAAGGATTTCTCTCG--TCAGCTTCCCG	1344
Db	41288	CTGGGCTCACTGCAACTCTTGCCCTCCGGGTCAAGGATTTCTTGCCCTCAAGCTCTCG	41222
QY	1345	AATAGCTGGGATTACAGGCGCATGCAACACATGC--CTAATTTTTGTATTTTTTAAGTAA	1402
Db	41228	AGTAGCTGGGACTACAGAGAGGTGCACCACTGCCACTAATTTGTATTTTTTAAGTAA	41166
QY	1403	GACAGAGTTTGCCATGTTGACCAAGGCTTGCTTGAATCTCTGACTTCAAGTGATCCACC	1462
Db	41168	GAAAGGGTTTACCATATTTGGTCAGTCTGGTCTTGAACTCTCTGACCTCAGGTATCCACC	41109
QY	1463	CACCTCAGGCTCCCAAAAGCACTGGAGTTTACAGGCATAGCAACGTCACGAGCTGTGTT	1522
Db	41108	CACCTCAGGCTCCCAAAAGTGTGGAGCTACAGGCATAGCAACCATGCCGCGCTGATAT	4104
QY	1523	CTGAGATCTGTATTTTGTCTGAAAGCTTCATTTCTATCTCTTATCAATTTGGAAGT	1582
Db	41048	TTAATCTTATTTATTTTATTTGCTAGCTATTAATCTTTCTATATGTCAACATTTCAAAGT	40988
QY	1583	AGTACACTTAAGTAAAGTTTTTAA CA 1608	
Db	40988	CATTTACAAATTTTAACTGCTTAAAA 40963	

RESULT 18  
 AP000085/c 100000 bp DNA linear PRI 25-MAY-2002  
 LOCUS AP000085 Homo sapiens genomic DNA of 21q22.1, App related, B2291C14-T1533  
 DEFINITION  
 region, segment 2/7.  
 ACCESSION AP000085  
 VERSION  
 KEYWORDS AP000085.1 GI:4730826  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominiidae; Homo.  
 REFERENCE  
 1 Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.  
 Homo sapiens 782,611bp genomic DNA of 21q22.1 App region  
 JOURNAL Published Only in Database (1999)  
 REFERENCE 2 (bases 1 to 100000)  
 Hattakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.  
 JOURNAL Direct Submission  
 TITLE Submitted (15-APR-1999) Mika Hirakawa, Japan science and Technology  
 JOURNAL Corporation (JST), Advanced Databases Department: 5-3, Yonbancho,  
 Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,  
 URL:htp://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,  
 Fax:81-3-5214-8470)  
 COMMENT This sequence is conducted by Kitasato University JST sequencing  
 laboratory as a JST sequencing team.  
 Principal Investigator:Yoshiyuki Sakaki Ph.D.  
 Phone: +81-3-5449-5622, Fax: +81-3-5449-5445  
 sakaki@ngc.ims.u-tokyo.ac.jp  
 Sub-leader: Tadayoshi Shiba Ph.D., Masahiro Hattori Ph.D. The  
 sequence is submitted by Human Genome Sequencing in ALIS project of  
 JST.  
 Japan Science and Technology Corporation (JST)  
 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive web site (<http://www-alis.tokyo.jst.go.jp/HGS/>) or send  
 email to [webmaster@ww-alis.tokyo.jst.go.jp](mailto:webmaster@ww-alis.tokyo.jst.go.jp).  
 Location/Qualifiers

	source	1. .100000	/organism="Homo sapiens"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:9606"	
			/chromosome="21"	
			/map="21q22.1"	
			53504. .53759	
			/standard_name="D1S265"	
			/note="AFM234was;Genethon Marker;The location is between each flanking site of PCR primers."	
			/db_xref="GDB:188463"	
ORIGIN				
Query Match	10.7%;	Score 214;	DB 8;	Length 100000;
Best Local Similarity	72.0%;	Pred. No. 3e-38;		
Matches 321;	Conservative 0;	Mismatches 120;	Indels 5;	Gaps 3;
OY	1167	TTTCAGGTGGAAATGTTGCAGAGAATAATTGTAATGTTTTCTCAGACTTTTATTT	12266	
Db	1309	TTCGAATGATRAATTTTGCCAGAGAAAACCTGSAATTTTTTTTTTTTTTTTTTTT	18500	
OY	1227	TTAATTTTTTGAACAAGAGCTCTCACTTTGTCAACCACCGCTGAGTACAGTGCCTGCT	12866	
Db	1849	TTTTTTTTTTTGAAGAAGAGCTGCTGCTGTCTCATCTCAGGCTGAGATGATGG-TGGCAT	17911	
OY	1287	CTGGGCTCACCTGCAACTCTTGCCCTCCAGGTTCAAAGGATTCCTCG--TCAGCTCCCC	13444	
Db	1790	CTGGGCTCACCTGCAACTCTGCCCTCCGGGTTCAAGGATTCCTGCTCAGCCTCGT	17311	
OY	1345	AATAGCTGGGATPACAGGCGCATCACACATGC--CTAATTTTTGTAATTTTAGTGA	14022	
Db	1730	AGTAGCTGGAGCTACAGAGAGCGTCCACACATCTCCAGCTAACCTTTTGATTTTAACTGA	16711	
OY	1403	GACAGAGTTTCGCAGATGTGACCAAGGCTTGCCCTTGAATCTCCGACTTCAGGTATCCA	14623	
Db	1670	GAAAGGGTTTACCATATTTGGTTCAGTCTGCTGTAACCTCTGAACCTCAGGTATGACCC	16111	
OY	1463	CACCTCAGCTCCCAAAAGCACTGGGATTAACAGGATGAGCCACCGTGCACGCTGTTT	15222	
Db	1610	CACCTCAGCTCCCAAAAGTGTGGAGACTACAGGATGAGCCACCAATGCCCGCTGGAAT	15511	
OY	1523	CTCAGATCCGTATTTGTTGTTGTTGAAGCCTCATTTCTATCTTTATTCATTTTGAAGT	15823	
Db	1550	TTAACTTATTTATTTATTTATTTATGCTGACTTATTAATGCTTTCTATATGCTAAGCATTC	14911	
OY	1583	AGTACACCTAAGTAAGTTTATTAACA 1608		
Db	1490	CATTTACAAATATTTAAGTCTTAADA 1465		
RESULT 19				
AP000137/c		100000 bp	DNA	linear PRI 08-JAN-2000
LOCUS				
DEFINITION	Homo sapiens genomic DNA, chromosome 21q21.2, L1S6-APP region,			
	clone B2291C14-R44F3, segment 2/10, complete sequence.			
ACCESSION	AP000137			
VERSION	AP000137.1 GI:4827103			
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 100000)			
AUTHORS	Hatori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y. Homo sapiens 911,949bp genomic DNA of 21q21.2 (REGION: L1S6-APP CLONE RANGE: B2291C14-R44F3) Published Only in Database (1999)			
JOURNAL	2 (bases 1 to 100000)			
REFERENCE	Hatori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y. Direct Submission			
TITLE				

## JOURNAL

Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555, Japan (E-mail:hattori@psc.riken.go.jp)  
URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-42-778-9923, Fax: 81-42-778-9924

## COMMENT

E. coli transposon insertion: The present data does not contain E. coli transposon sequences which integrated in the original/previous sequences. We determined the boundary between the insertion and genomic sequences experimentally, removed the insertion sequences, reconstituted the present data. The sequencing project is supported by Japan Science Technology Corporation (JST) and The Institute of Physical and Chemical Research (RIKEN).

## FEATURES

## source

1. 100000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q21.2"

## ORIGIN

Query Match 10.7%; Score 214; DB 8; Length 100000;  
Best Local Similarity 72.0%; Pred. No. 3e-38;  
Matches 321; Conservative 0; Mismatches 120; Indels 5; Gaps 3;

1167 TTTGAGTGTGAAATTTGTTCAAGAAAGTATTGATTTATTTTCAGATCTTTTATTT 1226  
1909 TTGAATGTGAAATTTGTTCAAGAAAGTATTGATTTATTTTCAGATCTTTTATTT 1850  
1227 TTAATTTTGTGAAACAGAGTCTCACTTTGTACACGAGCTGAGATGAGGCTGTGTT 1286  
1849 TTTTGTGTTTGTGAAAGAGAGTCTCGTCTGTACTGACGCTGAGTGCATTGG-TGCAT 1791  
1287 CTGGCTCACTGCAACCTCTGCTCCAGAGTTCAAGCAATTCCTG--TCAGTCTCCG 1344  
1790 CTGGCTCACTGCAACCTCTGCTCCAGAGTTCAAGCAATTCCTG--TCAGTCTCCG 1731  
1345 AATAGCTGGAGTTACAGCGCATGCAACCATGCG--CTAATTTTGTATTTTATGTA 1402  
1730 AGTACTGGGAGTACAGAGCGCTGCGACCATCCAGCACTTGTATTTTATGTA 1671  
1403 GACAGAGTTTGGCCATGTTGACAGAGCTTGCCTGAACCTGACCTGAGTATCCAC 1462  
1670 GAAAGGTTTACCAATTTGTCAGTCTGCTGCTGCAACTCTGACCTCAAGGATGAC 1611  
1463 CACCTCAGCTCCCAAGAGCACTGGATTACAGAGCATGAGCCACCGTGCACAGCTGTT 1522  
1610 CACCTCAGCTCCCAAGAGCTGGATTACAGAGCATGAGCCACCGTGCCTGGATAT 1551  
1523 CTGAGATCTGATTTGTTTCTGAAGCCTTCAATTTCTATCTTATTTCAATTTGAA 1582  
1550 TTAATCTTATTAATTTTATTTGCTACTATTAATGCTTTCTATATCTAAGCATCA 1491  
1583 AGTACACCTAAGTAAAGTTTATACA 1608  
1490 CATTTCATAATTAATTAAGCTTAAAA 1465

RESULT 20  
AP001693/C 340000 bp DNA linear PRI 21-MAY-2003

DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 37/105.

AP001693 AL163238 BA000005

VERSION AP001693.1 GI:7768672

## KEYWORDS

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.

REFERENCE 1 Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T.,

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

## PUBMED

## AUTHORS

```
/chromosome="21"
/map="21q21.2"
/clone="R266318, 5' partial"
/clone_lib="RPCT-11 BAC library"
/note="Accession No. AP001341"
63192. .121962
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.2"
/clone="CTD2289H10"
/clone_lib="CTD BAC library"
/note="Accession No. AP001340"
12143. .140060
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.2"
/clone="f39G8"
/clone_lib="CMF21 Fosmid library"
/note="Accession No. AP001348"
139705. .200913
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.2"
/clone="CTD2291C14"
/clone_lib="CTD BAC library"
/note="Accession No. AP000220"
200207. .289529
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.2"
/clone="pS672"
/clone_lib="CMF21 P1 library"
/note="Accession No. AP000221"
271109. .311125
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.2"
/clone="Q11L10"
/clone_lib="L121NCO2-Q Cosmid library"
/note="Accession No. AL109616"
310004. .>340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.2"
/clone="pS11, 3' partial"
/clone_lib="CMF21 P1 library"
/note="Accession No. AP000223"
2239. .2539
/note="AluSk"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(2607. .2851)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
3128. .3427
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(5176. .5395)
/note="MIR"
```

```
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
5492. .5603
/note="(TAGA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
5610. .5632
/note="(TCGA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
5633. 5654
/note="(ATG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
5655. .5716
/note="(TCGA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(5841. .5923)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(6425. .6667)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
7339. .7494
/note="L1ME3A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
7495. .7527
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
7528. 7775
/note="L1ME3A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(7867. .8094)
/note="MER20"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED

Query Match 10.7%; Score 214; DB 8; Length 340000;
Best Local Similarity 72.0%; Pred. No. 2.5e-38;
Matches 321; Conservative 0; Mismatches 120; Indels 5; Gaps 3;

QY 1167 TTTCAGTGTGAATGTTGTCAGAGAAAGTATGATTATGTTTCTCAATCTTTTATTT 1226
DB 241613 TTGAATTGATATATTTGTCCAGAGAAACCTTGATTTTTTTTTTTTTTTTTTTT 241554

QY 1227 TTAATTTTTTGAACAGAGTCTCACTTGTACCCAGGCTGGAGTACAGTGGCTGTGGT 1286
DB 241553 TTTTTTTTTTGAAGAGAGAGTCTCGCTCTGTCTACTCAAGCTGAGATGCAATGG-TCGCAAT 241495

QY 1287 CTCGGCTCACTGCAACTCTGCTCCAGGTTCAAGCAATCTCTG--TCAAGTCCCG 1344
DB 241494 CTCGGCTACTGCAACTCTGCTCTCTCGGTTCAAGCAATCTTCTGCTCAGCTCCG 241435

QY 1345 AATAGCTGGATTACAGGCGCATGCACCACTGC--CTAATTTTGTATTTTATAGTAA 1402
DB 241434 ACTAGCTGGACTACAGGAGGAGCTGTCACCACTCCAGACTTAACTTTTGTATTTTGTATA 241375

QY 1403 GACAGAGTTTGGCCANGTTGACAGGCTTGCTTGAACCTCGACTTCAGTGTATCCACC 1462
DB 241374 GAAAGGTTTACACNATTTGTGTAAGTGTGTCTGAACTCTGAACTCAGGTGATGACCC 241315

QY 1463 CACCTCAGCTCCCAAAGCATGAGATTACAGGCAATGACCAACCGTCCAGCCTGTGTT 1522
DB 241314 CACCTCAGCTCCCAAAGTGTGAGGACTACAGGCAATGACCAACATGCCGCTGGAATAT 241255

QY 1523 CTCAGATCTGTATTTGTTTCTGAAAGCTTCAATTTCTATCTTATTTATTTGGAAT 1582
DB 1523 CTCAGATCTGTATTTGTTTCTGAAAGCTTCAATTTCTATCTTATTTATTTGGAAT 1582
```

Db	241234	TTAACTATNTATTATTTTATTCGACGTATTAAGCTTGCTTGCATNTGCTAAGCATTCGAAGT	241193
Oy	1583	ACTAACCTTAAGTAGCTTTTAAACA	1608
LOCUS	AC110054/c		
DEFINITION	Homo sapiens chromosome 15, clone CTD-2116N17, complete sequence.		
ACCESSION	AC110054		
VERSION	AC110054.4	GI:20377046	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 164857)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
TITLE	Homo sapiens chromosome 15, clone CTD-2116N17		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 164857)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A., Camarata,U., Campioni,A., Chang,J., Chararo,B., Brown,A., Camarata,U., Campioni,A., Campiano,A., Chang,J., Chararo,B., Choepel,Y., Dearellano,K., Collins,S., Collymore,A., Cook,A., Cooke,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gaidyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meltrim,J., Menes,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnuppach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Teste,J., Theodore,J., Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliou,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 164857)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A., Camarata,U., Campioni,A., Chang,J., Chararo,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Farro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meltrim,J., Menes,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnuppach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Teste,J., Theodore,J., Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliou,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		

JOURNAL	Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	4 (bases 1 to 164857)
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Batra,N., Baetsen,V., Bloom,T., Boguslavsky,L., Bokhgalter,B., Brown,A., Cantarel,J., Campiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Deatrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-pierre,N., Haas,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karats,A., Kells,C., Lacroque,K., Lamazares,R., Landers,T., Lehoczkyl,J., Levine,R., Lindblad-Ton,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menous,L., Milova,T., Molina,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rossetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W., Young,G., Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On May 1, 2002 this sequence version replaced gi:19033659. All repeats were identified using RepeatMasker: Smt,A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
FEATURES	----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIRB Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information Center project name: L23190 Center clone name: 2116_N17 ----- Only the first 164,9 kilobases of this clone are being submitted. The remainder overlaps accession number AC100840 [WICGR project L21571].
SOURCE	Location/Qualifiers 1..164857 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="15" /map="15" /clone="CTD-2116N17" /clone_id="CTD1 Human BAC" complement(233..311) /rpt_family="MERSA" 644..664 /rpt_family="AT_rich" complement(665..962) /rpt_family="AluY" 963..980 /rpt_family="AT_rich" complement(1116..1519) /rpt_family="AluXs" complement(2545..2642) /rpt_family="AluO" 2845..3183 /rpt_family="L2" 3108..3331 /rpt_family="(A)n" complement(3384..3686) /rpt_family="AluSg" complement(3691..3848)

```
repeat_region      /rpt_family="L2"
                    3849..4144
                    /rpt_family="AluDb"
repeat_region      complement(4145..4298)
                    /rpt_family="L2"
repeat_region      complement(4326..4545)
                    /rpt_family="L2"
repeat_region      4918..5042
                    /rpt_family="MER85"
repeat_region      complement(5204..5319)
                    /rpt_family="AluY"
repeat_region      complement(5334..5640)
                    /rpt_family="AluSg"
repeat_region      5896..6188
                    /rpt_family="AluSx"
repeat_region      6298..6332
                    /rpt_family="AT_rich"
repeat_region      complement(6346..6633)
                    /rpt_family="MER1B"
repeat_region      complement(6664..6926)
                    /rpt_family="L1MB"
repeat_region      complement(7320..7626)
                    /rpt_family="AluSc"
repeat_region      complement(8043..8352)
                    /rpt_family="AluY"
repeat_region      complement(8660..8962)
                    /rpt_family="AluSg"
repeat_region      9256..9442
                    /rpt_family="AluSg"
repeat_region      9443..9504
                    /rpt_family="AluSg"
repeat_region      9505..9539
                    /rpt_family="AluSg"
repeat_region      complement(9567..10720)
                    /rpt_family="L1PB3"
                    10683..10783
                    /note="single clone coverage"
unsure
repeat_region      complement(11072..11376)
                    /rpt_family="AluSx"
repeat_region      11660..11671
                    /rpt_family="(CA)n"
repeat_region      11672..11706
                    /rpt_family="(CA)n"
repeat_region      11737..11764
                    /rpt_family="(CA)n"
repeat_region      complement(11922..12167)
                    /rpt_family="AT_rich"
repeat_region      complement(12249..12438)
                    /rpt_family="HML1"
repeat_region      12445..12977
                    /rpt_family="L3"
repeat_region      complement(12984..13117)
                    /rpt_family="MIR3"
repeat_region      complement(13565..13678)
                    /rpt_family="L2"
repeat_region      15124..15145
                    /rpt_family="AT_rich"
repeat_region      15495..15524
                    /rpt_family="AT_rich"
repeat_region      15666..16484
                    /rpt_family="L1MB6"
repeat_region      16487..16620
                    /rpt_family="AluSg/x"

Query Match      10.7%; Score 213.8; DB 8; Length 164857;
Best Local Similarity 79.1%; Pred. No. 3.1e-38;
Matches 292; Conservative 0; Mismatches 72; Indels 5; Gaps 3;
```

```
QY      1182 TGTTCAGAGATTTGATTAATTTTCAGATCTTTTATTTTATTTTATTTTGA 1241
DB      30813 TGTTCAGAGAGGTTTGTGATGATTTTCATTTTATTTTATTTTATTTTGA 30754
QY      1242 CAGAGTCTGCTTTGTACCCAGGCTGAGTACAGTGGCTGGCTCGGCTCAGCA 1301
```

```
Db      30753 CGAGTCTTGTCTGTGTGCCCCAGGCTGAGTGCAGTGG-TGCATTTAGCTCAGTGC 30695
QY      1302 CCTTCAGCTCCAGATTTCAAGCATTTCTCTG--TCAAGCTTCCGAAATAGCTGGATTAC 1359
DB      30694 CCTTCAGCTCCAGGTTCAAGCATTTCTCTGCTCAGCTCCGAGTATAGTGGATTAC 30635
QY      1360 AGGCGATGACCAACCATGC--CTAATTTTGTATTTTATTTAGTACAGACAGTTTCGCCA 1417
DB      30634 AGGTGATATTACACAGCTCCAGCTAATTTTGTATTTTATTTAGACAGACAGGTTTTCGCCA 30575
QY      1418 TGTTCACAGGCTTCCCTTGAATCTTCAGTTCAGTGTATCCACCACTGAGCTGCCA 1477
DB      30574 TGTTCACAGGCTTCCCTTGAATCTTCAGTTCAGTGTATCCACCACTGAGCTGCCA 30515
QY      1478 AAGCAGTGGATTCAGGATGAGCATGAGCACCCTGAGCTGTTTCTCAGATTCCTGATTT 1537
DB      30514 AAGTCTGGATTCAGGATGAGCATGAGCACCACCATGCCCCGCTCATGATGCTATTTTATTTT 30455
QY      1538 TGTTCCTGA 1546
DB      30454 TATTTTGA 30446

RESULT 22
AR659559/c      72048 bp      DNA      linear      PAT 13-JUN-2005
LOCUS      AR659559/c
DEFINITION      Sequence 600 from patent US 6900016.
ACCESSION      AR659559
VERSION      AR659559.1 GI:67595583
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 72048)
AUTHORS      Venter,J.C., Zhang,J.N., Liu,X., Rowe,W., Crawchik,A., Kalush,F.,
              Naik,A., Subramanian,G. and Woodage,T.
TITLE      Polymorphisms in known genes associated with inflammatory
              autoimmune disease, methods of detection and uses thereof
JOURNAL      Patent: US 6900016-A 600 31-MAY-2005;
              Applera Corporation; Norwalk, CT
FEATURES
source      location/Qualifiers
              1..72048
              /organism="unknown"
              /mol_type="genomic DNA"
ORIGIN
Query Match      10.7%; Score 213.6; DB 6; Length 72048;
Best Local Similarity 74.0%; Pred. No. 3.9e-38;
Matches 311; Conservative 0; Mismatches 104; Indels 5; Gaps 3;
```

```
QY      1208 TTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTCCAGGCT 1267
DB      64752 TACTGTATTTCTTTTCTTTTATTTTGTAGACAAAGTCTCTCTGTCCACAGGCT 64653
QY      1268 GGAGTACAGTGGCTGTGATCTCGGCTCACTGCAACCTCTGCTCCAGATTCAAGCAT 1327
DB      64692 GGAGTACAGTGGCT -ACGATCTCAGCTCACTGCAACCTCTGCTCTCAGATTCAAGCAT 64634
QY      1328 CTCCTG--TCAGCTTCCGATATGCTGGATTTACAGGGCATGCACCA--TGCCTTA 1383
DB      64633 CTCCTGCTTACCTCCGAGTACGCTGGATTTACAGGGCATGCACCACTGCTGCTAC 64574
QY      1384 TTTTGTATTTTATTTAGAGACAGAGTTTGCGCATGTTGACAGGCTTGTGCTTGAATCC 1443
DB      64573 TTTTGTATTTTATTTAGAGACAGAGTTTGCGCATGTTGCTGCTGCTTGAATCC 64514
QY      1444 TGACTTCAGAGTATCCACCACTGAGCTCCCAAGCACTGGATTTACAGGATGAGCC 1503
DB      64513 TGACTTCAGAGTATCCACCACTGAGCTCCCAAGGCTGGATTTACAGGATGAGACC 64454
QY      1504 ACCGTGCCAGCTGTTTCTCAGATCCGTATTTTGTTCGAAAGCTTCAATTTCTATCT 1563
```

Db 64453 ACCGTCGCCGCGCATGCTGTAATTTCTTTGTAATTAATTCCTAAGTAATTTATCT 64394

QY 1564 TCTTATTCATTTTGAAGTAGTACACTAGTAGAGTTTATCAATCAATCTTTGG 1623

Db 64393 TTTTAATCCATATGTAAATGGAGTTTAAATAATTAATGTTTGGATGTTTATTTCTAG 64334

RESULT 23

AR659643/c AR659643 72048 bp DNA linear PAT 13-JUN-2005

DEFINITION Sequence 684 from patent US 6900016.

ACCESSION AR659643

VERSION AR659643.1 GI:67595683

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 72048)

AUTHORS Venter,J.C., Zhang,J.N., Liu,X., Rowe,W., Cravchik,A., Kalush,F., Naik,A., Subramanian,G. and Woodage,T.

TITLE Polymorphisms in known genes associated with inflammatory autoimmune disease, methods of detection and uses thereof

JOURNAL Patent: US 6900016-A 684 31-MAY-2005;

FEATURES

source 1..72048

Location/Qualifiers

1..72048

/organism="unknown"

/mol\_type="genomic DNA"

ORIGIN

Query Match 10.7%; Score 213.6; DB 6; Length 72048;

Best Local Similarity 74.0%; Pired. No. 3.9e-38;

Matches 311; Conservative 0; Mismatches 104; Indels 5; Gaps 3;

QY 1208 TTCTCAGATCTTTTATTTTATTTTATTTTGAAGAGAGCTCCTTGTCAACCCAGGCT 1267

Db 64752 TACTTGATTTCTTTTCTTTTCTTTTGTGACAGAGCTCTCTGTGACCCAGGCT 64693

QY 1268 GGAAGTACAGTGGTGTGCTGCTGCTGCTGCACTCTGCTCTCCAGGTTCAAGGANT 1327

Db 64692 GGAAGTACAGTGGTGTGCTGCTGCTGCTGCACTCTGCTCTCCAGGTTCAAGGANT 64634

QY 1328 CTCCTG--TCAGCTTCCCGAATAGCTGGATTCAGAGCGCATGCACACA--TGCTTA 1383

Db 64633 CTCCTGCTTACGCTCCCGAGTAGCTGGATTCAGAGCGCATGCACACA--TGCTTA 64574

QY 1384 TTTTGTATTTTATAGTAGAGACAGAGTTTCAGCATGTGACAGGCTTGGAATCTC 1443

Db 64573 TTTTGTATTTTATAGTAGAGAGAGGTTTCAGCATGTGAGTGGTGTGGAATCTC 64514

QY 1444 TGACTTCAGGTATTCACCACTCAGCTCCCAAGCACTGGGATTTACAGGATGAGCC 1503

Db 64513 TGACTTCAGGTATTCACCACTCAGCTCCCAAGCACTGGGATTTACAGGATGAGCC 64454

QY 1504 ACCGTCGCCGCGCATGCTGTAATTTCTTTGTAATTAATTCCTAAGTAATTTATCT 1563

Db 64453 ACCGTCGCCGCGCATGCTGTAATTTCTTTGTAATTAATTCCTAAGTAATTTATCT 64394

QY 1564 TCTTATTCATTTTGAAGTAGTACACTAGTAGAGTTTATCAATCAATCTTTGG 1623

Db 64393 TTTTAATCCATATGTAAATGGAGTTTAAATAATTAATGTTTGGATGTTTATTTCTAG 64334

RESULT 24

AC068969/c AC068969 168751 bp DNA linear HTG 07-MAY-2001

LOCUS Homo sapiens chromosome 15 clone RP11-185A14 map 15, WORKING DRAFT

DEFINITION

SEQUENCE 12 unoriented pieces.

ACCESSION AC068969

VERSION AC068969.3 GI:13959305

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 15, clone RP11-185A14

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 168751)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baidin,J., Bara,N., Bastien,V., Bede,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,S., Ferreira,P., Fitzhugh,W., Gage,D., Galsgan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Liu,X., Locke,K., MacDonald,P., Margulis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrim,J., Menes,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,I., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talmas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Ttigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL Submitted (14-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 7, 2001 this sequence version replaced gi:8247910.

All repeats were identified using RepeatMasker:

Smit,A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/BM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9836

Center clone name: 185.A.14

----- Summary Statistics

Sequencing vector: M13; M77815; 91% of reads

Sequencing vector: Plasmid; n/a; 9% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960721

Consensus quality: 162617 bases at least Q40

Consensus quality: 165848 bases at least Q30

Consensus quality: 167104 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 167651; sum-of-contigs

Quality coverage: 5.4 in Q20 bases; agarose-fp

Quality coverage: 5.5 in Q20 ba.

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1

5645: contig of 5645 bp in length

5646: gap of 100 bp

5746: contig of 1172 bp in length

5746: gap of 100 bp

6918: gap of 100 bp

7018: contig of 1103 bp in length

8320: contig of 1303 bp in length

8321: gap of 100 bp

8420: gap of 100 bp

8421: contig of 2133 bp in length

10553: contig of 2133 bp in length

10554: gap of 100 bp

10654: contig of 2890 bp in length

	*	13544	13643: gap of 100 bp
	*	13644	19241: contig of 5598 bp in length
	*	19242	19341: gap of 100 bp
	*	19342	35264: contig of 15923 bp in length
	*	35265	35364: gap of 100 bp
	*	35365	68938: contig of 33574 bp in length
	*	68939	69038: gap of 100 bp
	*	69039	90272: contig of 21234 bp in length
	*	90273	90372: gap of 100 bp
	*	90373	113565: contig of 23193 bp in length
	*	113566	113665: gap of 100 bp
	*	113666	146203: contig of 35358 bp in length
	*	146204	146303: gap of 100 bp
	*	146304	168751: contig of 22448 bp in length
FEATURES			Location/Qualifiers
source			1..168751
			/organism="Homo sapiens"
			/mol_type="genomic DNA"
			/db_xref="taxon:9606"
			/chromosome="15"
			/map="15"
			/clone="RP11-185A14"
			/clone_id="RPC1-11 Human Male BAC"
misc_feature			1..5645
			/note="assembly_fragment"
			clone_end:SP6
gap			vector_side:left"
			5646..5745
			/estimated_length=100
misc_feature			5746..6917
			/note="assembly_fragment"
gap			6918..7017
			/estimated_length=100
misc_feature			7018..8320
			/note="assembly_fragment"
gap			8321..8420
			/estimated_length=100
misc_feature			8421..10553
			/note="assembly_fragment"
gap			10554..10653
			/estimated_length=100
misc_feature			10654..13543
			/note="assembly_fragment"
gap			13544..13643
			/estimated_length=100
misc_feature			13644..19241
			/note="assembly_fragment"
gap			19242..19341
			/estimated_length=100
misc_feature			19342..35264
			/note="assembly_fragment"
gap			35265..35364
			/estimated_length=100
misc_feature			35365..68938
			/note="assembly_fragment"
gap			68939..69038
			/estimated_length=100
misc_feature			69039..90272
			/note="assembly_fragment"
gap			90273..90372
			/estimated_length=100
misc_feature			90373..113565
			/note="assembly_fragment"
gap			113566..113665
			/estimated_length=100
misc_feature			113666..146203
			/note="assembly_fragment"
gap			146204..146303
			/estimated_length=100
misc_feature			146304..168751
			/note="assembly_fragment"
misc_feature			vector_side:right"

Query Match	10.7%;	Score 213.6;	DB 14;	Length 168751;
Best Local Similarity	74.0%;	Pred. No. 3.5e-38;	Mismatches 104;	Indels 5; Gaps 3;
Matches 311;	Conservative 0;			
Oy	1208	TTCTCAATCATCTTATTATTTTAAATTTTGTGAAAACAGAGTCCTACATTGTCACCCAGGCT	1267	
Dd	47984	TACTTGATTTCTTTTTCTTTTTTTTTTTTGAAGCAGAAGCTCTCTCTGTGCACCAGGCT	47922	
Oy	1268	GGAATACAGTGGCTGTGTGCTCTCGGCTCACATGCAACTCTGCTCTCCACAGTTCAAAGCATT	1327	
Dd	47924	GGAATGACAGTGGC-ACGATCTCAGCTCACATGCAAACTCTGCTCTCAGTTCAAAGCATT	4786	
Oy	1328	CTCTGT--TAGCTTCCCCAATAGCTGGGATTTACAGGCGCATGACACACA--TGCTTA	1383	
Dd	47865	CTCTGTCTTAAAGCTCTCCCGAGTGTGGGATTAACAGGTGTGCAGCACACACCTGGGCTAC	4780	
Oy	1384	TTTTTGATTTTATTTAGTACAGACAGATTTGGCATTTGTGACCAAGGCTTGGCTTAACCTCC	1443	
Dd	47805	TTTTTTGATTTTATTTAGTACAGGAGATTTTGCCATGTTGGTGTGACGGCTGTGTAATCCC	4774	
Oy	1444	TGACTTCAGGTGATTCACCACTCAGCTCAGCTCCCAAGACACTGGGATTAACAGCATGAGCC	1503	
Dd	47745	TGACCTCAGGTGATTCACCACTCAGCTCAGCTCCCAAGATGTGGGATTAACAGCATGAGAAC	4768	
Oy	1504	ACCGTGGCCAGCCCTGTTTTCTCAGATCTCTGATTTGTTTCTGAAAGCCTTGATTTCTATCT	1563	
Dd	47685	ACCGTGGCCGCCAGCTGTTTTCTGATTTCTTTTGTGTAATAATTAATTCATGATTTATTTCT	4762	
Oy	1564	TCTTATTCATTTTGAAGTACACCTTAAGTAAAGTTTAAAAACAATCAATAATCTTTGG	1623	
Dd	47625	TTTTTAATCTTATTTGTAATGGGATTTTAAAAAATTAATGTTTGTGATTTGTTATTTCTAG	4756	
RESULT 25				
AC067805/c				
LOCUS	AC067805	170883 bp	DNA	linear PRI 25-NOV-2001
DEFINITION	Homo sapiens chromosome 15, clone RP11-63A23, complete sequence.			
ACCESSION	AC067805			
VERSION	AC067805.9	GI:17061076		
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
	Hominidae; Homo.			
REFERENCE	1 (bases 1 to 170883)			
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.			
TITLE	Homo sapiens chromosome 15, clone RP11-63A23			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 170883)			
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,			
	Anderson,S., Baldwin,T., Barna,N., Bastien,V., Bedalov,F.,			
	Boguslavsky,I., Bouckgeyser,B., Brown,A., Burkett,G.,			
	Campopiano,A., Castle,A., Choapel,Y., Colangelo,M., Collins,S.,			
	Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,U.S.,			
	Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,			
	Gargan,J., Gardina,S., Ginde,S., Goeyette,M., Graham,L.,			
	Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,			
	Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,			
	Klein,J., Lahocque,K., Lamaizares,R., Landers,T., Lehoczyk,U.,			
	Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,			
	McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McNeeters,R.,			
	Meldrum,T., Meneses,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,			
	Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,			
	O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,			
	Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,			
	Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,			
	Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,			
	Tesfaye,S., Theodore,T., Tirrell,A., Travers,X., Trifilio,U.,			
	Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,			
	Young,G., Zainoun,T., Zimmer,A. and Zody,M.			



TITLE Direct Submission  
JOURNAL Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 170883)  
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand, P., Hagen, N.,  
Hagos, B., Heathford, A., Horton, L., Hulme, W., Illier, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,  
Lamasares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Margis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 23, 2001 this sequence version replaced gi:16554401.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L9837  
Center clone name: 63\_A\_23  
-----

FEATURES  
source location/Qualifiers  
1..170883  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone="RP11-63A23"  
/clone\_11b="RP11-11 Human Male BAC"  
complement(427..489)  
repeat\_region  
/rpt\_family="L2"  
520..594  
/rpt\_family="MIR"  
895..1068  
/rpt\_family="MIR"  
complement(1233..1527)  
repeat\_region  
/rpt\_family="AluSk"  
complement(1707..1826)  
repeat\_region  
/rpt\_family="MER112"  
1847..1918  
/rpt\_family="L2"  
2628..2666  
/rpt\_family="L2"  
complement(3097..3216)  
repeat\_region  
/rpt\_family="MIR"  
complement(3246..3375)  
repeat\_region  
/rpt\_family="L2"  
3414..3652  
/rpt\_family="MER20"  
3811..4004  
repeat\_region  
/rpt\_family="MIR"

repeat\_region 4208..4301  
/rpt\_family="MIR"  
complement(4895..5001)  
repeat\_region  
/rpt\_family="MIR3"  
complement(5600..5751)  
repeat\_region  
/rpt\_family="MERS5B"  
7380..7697  
/rpt\_family="L2"  
7965..8075  
/rpt\_family="L2"  
9003..9162  
repeat\_region  
/rpt\_family="MIR"  
9475..9563  
repeat\_region  
/rpt\_family="MER91C"  
10157..10454  
/rpt\_family="AluSk"  
10537..10563  
repeat\_region  
/rpt\_family="L2"  
complement(10705..11175)  
/rpt\_family="MIR"  
complement(11404..11516)  
repeat\_region  
/rpt\_family="MIR"  
complement(12012..12321)  
/rpt\_family="Alu"  
12452..12757  
repeat\_region  
/rpt\_family="AluDb"  
complement(12767..12872)  
/rpt\_family="L1MB3A"  
complement(12988..13085)  
repeat\_region  
/rpt\_family="L2"  
13670..13826  
/rpt\_family="L1MB3"  
14301..14334  
repeat\_region  
/rpt\_family="L2"  
14361..14652  
/rpt\_family="AluDo"  
14654..14678  
repeat\_region  
/rpt\_family="AT-rich"  
14787..14997  
/rpt\_family="MIR"  
15459..15563  
repeat\_region  
/rpt\_family="L2"  
complement(15651..15701)  
/rpt\_family="AluU"  
complement(15702..16075)  
repeat\_region  
/rpt\_family="THE1C"  
complement(16076..16202)  
/rpt\_family="AluU"  
complement(16249..16331)  
repeat\_region  
/rpt\_family="AluU"  
complement(16668..17193)  
/rpt\_family="MER77"  
complement(17350..17613)  
repeat\_region  
/rpt\_family="AluSk"  
18923..19228  
/rpt\_family="AluY"  
complement(19290..19476)  
repeat\_region  
/rpt\_family="L2"  
19477..19781  
repeat\_region  
/rpt\_family="AluSk"  
complement(19782..20084)  
/rpt\_family="L2"  
21246..21273  
/rpt\_family="L2"  
21770..21894  
repeat\_region  
/rpt\_family="MIR"  
complement(22641..22815)  
repeat\_region  
/rpt\_family="MIR"  
complement(24142..24281)  
/rpt\_family="MIR"  
complement(25199..25875)  
/rpt\_family="MER7C"  
25977..26089  
repeat\_region



		/note="assembly_fragment clone_end:T7 vector_side:left" 1..36892
		/note="clone overlaps with GenBank Accession Number AC091717 clone RP43-184N1 (center project name aoz)" 68460..68559
	gap	/estimated_length=unknown 68560..152280
	misc_feature	/note="assembly_fragment" 131176..175752
	misc_feature	/note="clone overlaps with GenBank Accession Number AC091716 clone RP43-147G10 (center project name ceb)" 152281..152380
	gap	/estimated_length=unknown 152381..166826
	misc_feature	/note="assembly_fragment" 166827..166926
	gap	/estimated_length=unknown 166927..175752
	misc_feature	/note="assembly_fragment clone_end:Sp6 vector_side:right"
	ORIGIN	
	Query Match	10.7%; Score 213.4; DB 14; length 175752;
	Best Local Similarity	61.4%; Pred. No.3.8e-38;
	Matches 412; Conservative	0; Mismatches 251; Indels 8; Gaps 4;
OY	1205 GTTTCCTAGATCTTTTATTATTTTATTTTTTGAAACAGAGCTCATCTTGTGCACCAG	1264
Db	127196 GCTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATAACCGAAGCTGCTCTGCATCAG	127255
OY	1285 GCTGAGTAGACGTGCTGCTGCTGCTCACTGCACCTCTGCCCTCCAGTTCAAGCG	1324
Db	127256 GCTGAGTAGACGTGAC-GGGGTCTGGGCTCACCTGCACAACCTCTGCCAGGTTCAAGCG	127314
OY	1335 ATTCTCCG--TCACCTTCCCCGAATAGCTGGATTACAGGGCATG----CACCACTAG	1378
Db	127315 ATTCCTCCGCTCACGCTCCCAGAATAGCTGGATTACAGGCGTCTGCATATACACCCA	127374
OY	1379 CCTAATTTTTGTATTTTTGTATAGAGACAGAGTTTCCGATGTGACAGGCTTGCCTTCA	1438
Db	127375 GCATATTTTGTATTTTATTTATAGTAGAGCGGGTTTCAACATGTGGCAGGCTGTCTCCA	127434
OY	1439 ACTCTGACTTCAAGTGATCAACCCACTCACTCACTCCCAAAGCATGGGATTACAGGCAT	1498
Db	127435 ACTCTGACTTCAAGTGATCTCTGCTCTGCTCCAGAGTGTGGGATTACAGGCAT	127494
OY	1499 GAGCAACGTGCCAGCCCTGTTTTCTCAATCTGTATTGTTTGAAGCCCTTCAATTTC	1558
Db	127495 GAGCACTGCGCCGCCGCCAGCTCTGTTCTTTTTTTTAAATTGAATAAATCTGT	127554
OY	1559 TATCTTCTATTCATTTTGGAGAGTAGACACCTAAGTAAGTTTAAACATCAATAATC	1618
Db	127555 GTTGTGTTCTTTGA-TTGAANAATCTGAGGTTCATATGTATTTCTTATAACGTGA	127613
OY	1619 TTGGAAAATTCCTCGGTCTCTTCTTATTTCTTCAAAAAATATGTTCAAGTAGCTGAATG	1678
Db	127614 TTTCTGTGTTTGTAGTATTCCTTTTAAATTTTATAGAAAATCCTTATTTTGTAGATACG	127673
OY	1679 TTATGTTCTTCAATATTAATTCATTTCTGTATCTGATCTAGAAATTTATCTCAGSCCAATGTT	1738
Db	127674 TTTTTCATAGCAATATGAATTAAGTATTTTCAATTTTTTCTCTCTCCCGTG	127733
OY	1739 ATGGAATGCTTCACTTCTGTCACTCAAGTTCTGAGTCTCTTATTTCACTGAAGCTA	1798
Db	127734 TTCATATTTTATGTTATCAATTAACCTAAGTTGCTTTATTACTATTAAATTTTTTTTG	127793
OY	1799 AGTGCTATTGAATTAAGAGCTTGTAACAGATTTCTTCCAAATAGCTTATCTTTG	1858
Db	127794 TACTTCTCTTGCCTTTGTAGCTTTAAATAATCTTTTGACACCTAGTATTTATATCA	127853

QY 1859 ACTGCATGCCA 1869  
|||  
Db 127854 TCTTCCCCCCA 127864

RESULT	27
AC146040/c	
LOCUS	AC146040
DEFINITION	Pan troglodytes BAC clone RP43-97N16 from 7, complete sequence.
ACCESSION	AC146040
VERSION	AC146040.2 GI:40254142
KEYWORDS	HTG.
SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Pan troglodytes

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (bases 1 to 175808)	Hackins, R., Kozlowski, A. and Bielicki, J.	The sequence of Pan troglodytes BAC clone RP43-97N16	Unpublished (2001)	2 (bases 1 to 175808)

TITLE	REFERENCE
Sequencing of Pan troglodytes Unpublished (2001)	3 (bases 1 to 175808)

**JOURNAL**  
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA

**TITLE** Direct Submission  
**JOURNAL** Submitted (20-DEC-2003) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA  
**REFERENCE** 5 (bases 1 to 175808)

**TITLE** Direct Submission  
**JOURNAL** Submitted (03-JAN-2004) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
**COMMENT** On Dec 20, 2003 this sequence version replaced gi:33386592.

Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@wustl.wustl.edu](mailto:submissions@wustl.wustl.edu)  
 ----- Summary Statistics -----  
 Center project name: C\_PT097N16

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

**SOURCE INFORMATION:**  
The RPcI-43 BAC library has been constructed by Chung-li Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The



```
AL590822      179836 bp      DNA      linear      PRI 18-MAY-2005
LOCUS      Human DNA sequence from clone RP11-181G12 on chromosome 1 contains
DEFINITION the 3' end of the PRKCZ gene for protein kinase C zeta, a novel
            gene, a novel gene (FLJ31031), the SKI gene for v-SKI sarcoma viral
            oncogene homolog (avian) and five CpG islands, complete sequence.
ACCESSION   AL590822
VERSION     AL590822.36  GI:32251606
KEYWORDS    HTG; FLJ1031; PRKCZ; SKI.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE   1 (bases 1 to 179836)
AUTHORS     Hall, R.
TITLE       Direct Submision
JOURNAL     Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
            On Jun 25, 2003 this sequence version replaced gi:31071456.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Ch1
            RP11-181G12 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACE3.6
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vegas@sanger.ac.uk
            -----
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Location/Qualifiers
                1. 179836
                   /organism="Homo sapiens"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:9606"
                   /chromosome="1"
                   /clone="RP11-181G12"
                   /clone_1fb="RPCI-11.1"
            1
            /note="clone left end: RP11-181G12"
            join(AL391845.49:90455..90686,AL391845.49:95426..95547,
            AL391845.49:96469..96558,AL391845.49:99526..99576,
            AL645703.18:44363..44448,3076..3207,4893..4974,7738..7790,
            9656..9844,14861..14958,28384..28470,30921..31056
            31167..31254,32763..32882,33620..33699,34090..34179,
            43449..43564,43788..44261)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-001"
            join(AL391845.49:90455..90686,AL391845.49:95426..95547,
            AL391845.49:96469..96558,AL391845.49:99526..99576,
            AL645703.18:44363..44448,3076..3207,4893..4974,7738..7790,
            9656..9844,14861..14958,28384..28470,30921..31056,
            31167..31254,32763..32882,33620..33699,34090..34179,
            43449..43564,43788..44261)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-001"

misc_feature
            1
            /note="clone left end: RP11-181G12"
            join(AL391845.49:90455..90686,AL391845.49:95426..95547,
            AL391845.49:96469..96558,AL391845.49:99526..99576,
            AL645703.18:44363..44448,3076..3207,4893..4974,7738..7790,
            9656..9844,14861..14958,28384..28470,30921..31056,
            31167..31254,32763..32882,33620..33699,34090..34179,
            43449..43564,43788..44261)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-001"

gene
            1
            /note="clone left end: RP11-181G12"
            join(AL391845.49:90455..90686,AL391845.49:95426..95547,
            AL391845.49:96469..96558,AL391845.49:99526..99576,
            AL645703.18:44363..44448,3076..3207,4893..4974,7738..7790,
            9656..9844,14861..14958,28384..28470,30921..31056,
            31167..31254,32763..32882,33620..33699,34090..34179,
            43449..43564,43788..44261)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-001"

CDS
            1
            /product="protein kinase C, zeta"
            /note="match: cDNAs: Em:BC008058.1 Em:BC008058.2
            Em:BC014270.1 Em:BC014270.2 Em:BC007082.1 Em:LI4283.1"
            join(AL391845.49:90616..90686,AL391845.49:95426..95547,
            AL391845.49:96469..96558,AL391845.49:99526..99576,
            AL645703.18:44363..44448,3076..3207,4893..4974,7738..7790,
            9656..9844,14861..14958,28384..28470,30921..31056,
            31167..31254,32763..32882,33620..33699,34090..34179,
            43449..43564,43788..43875)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-001"
            /standard_name="ORTHOMPO0000001368"
            /note="match: proteins: Sw:019111 Sw:P09217 Sw:002956
            Sw:005513 Tr:ABP35745 Tr:Q9JLP6"
            /codon_start=1
            /product="protein kinase C, zeta"
            /protein_id="CAI1484.1"
            /db_xref="GI:55961168"
            /translation="MPSRTGPMKSGSGRVLKAHYGDIPTSVDAATTFEELCEEV
            RDMCRHQOHPLTLKWDSEDPCTVSSOMLEAEFRLAROCRDGLIHHVPSRPEO
            EGLPCEGDKSIYRRGARMRKTLRANHLFOAKRFNRATOCGSEBIWGLARCYR
            CINCKLVKRCHEGLVPLTCRKMDSVMPSEPPVDKXEDADLSEETDGIAYISS
            RKHDSIKDSDSLKVIDMDGDKISQGLQDPDLIVIGSGYAKYLVLRKNDQ
            IYAMKLVKKELVHDEDDIDWQTEKRVFEQASNPFLVIGSCFQTSRLFLVIEYV
            GADLVFMQQRKLPSEHARFYAEICIALNPFGHLYRDKLVNLDADGHKL
            TDYCGKGLRPGDTSTFCGTPVYIAEILRGHRYGSPVMDLGVLMFEMAGRSF
            FDIITDNPMTEDETLFVYLAKPRIRPRISYKASHYLRKGFANQPKRIGCRPQTG
            FDIKSHAFPSISIDWDLKFKQALPPFPQPLTDDGLNFTQFSEVQLTPDEDA
            IKRIDQSEFEFEYINPLLSTESV"
            join(AL391845.49:113629..113914,AL645703.18:44363..44448,
            3076..3207,4893..4974,7738..7790,9656..9844,14861..14958,
            28384..28470)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-009"
            join(AL391845.49:113629..113914,AL645703.18:44363..44448,
            3076..3207,4893..4974,7738..7790,9656..9844,14861..14958,
            28384..28470)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-009"
            /product="protein kinase C, zeta"
            /note="match: ESTs: Em:AL536613.2 Em:B1550905.1"
            join(AL645703.18:44091..44187,AL645703.18:44363..44448,
            3076..3207,4455..4377,4893..4974,7738..7790,9656..9645)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-011"
            join(AL645703.18:44091..44187,AL645703.18:44363..44448,
            3076..3207,4455..4377,4893..4974,7738..7790,9656..9645)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-010"
            join(AL645703.18:20755..20918,AL645703.18:44363..44448,
            3076..3207,4893..4974,7738..7790,9656..9644)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-010"
            /product="protein kinase C, zeta"
            /note="match: ESTs: Em:CD25E012.1"
            join(AL645703.18:20755..20918,AL645703.18:44363..44448,
            3076..3207,4893..4974,7738..7790,9656..9644)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-010"
            /product="protein kinase C, zeta"
            /note="match: ESTs: Em:AL044351.1 Em:B1560781.1"
            join(AL645703.18:13817..13996,AL645703.18:44363..44448,
            3076..3207,4893..4974,7738..7790,9656..96736)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-003"
            join(AL645703.18:13817..13996,AL645703.18:44363..44448,
            3076..3207,4893..4974,7738..7790,9656..96736)
            /gene="PRKCZ"
            /locus_tag="RP11-181G12.1-003"
            /product="protein kinase C, zeta"
            /note="match: ESTs: Em:AL54576.2 Em:AU137858.1
            Em:AU138321.1 Em:AU138397.1 Em:AU13706.1 Em:AU138889.1
            Em:BG571791.1"
            /locus_tag="RP11-181G12.1-001"
```

gene  
join(AL645703.18:42826..43101,AL645703.18:44363..44448,  
3076..3207,4893..4974,7738..>7790)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-004"  
join(AL645703.18:42826..43101,AL645703.18:44363..44448,  
3076..3207,4893..4974,7738..>7790)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-004"  
/product="protein kinase C, zeta"  
/note="match: ESTs: Em:BE517476.1 Em:RP125400.1"  
join(AL91845.49:90470..90686,AL91845.49:95426..95547,  
AL91845.49:99526..99576,AL645703.18:44363..44448,  
3076..3207,4893..>4974)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-007"  
join(AL91845.49:90470..90686,AL91845.49:95426..95547,  
AL91845.49:99526..99576,AL645703.18:44363..44448,  
3076..3207,4893..>4974)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-007"  
/product="protein kinase C, zeta"  
/note="match: ESTs: Em:BG329855.1"  
join(AL645703.18:43914..44187,AL645703.18:44363..44448,  
3076..3207,4893..>4964)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-006"  
join(AL645703.18:43914..44187,AL645703.18:44363..44448,  
3076..3207,4893..>4964)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-006"  
/product="protein kinase C, zeta"  
/note="match: ESTs: Em:BG898174.1"  
join(3317..3450,4893..4974,7738..7790,9656..9844,  
14861..14988,28384..28470,30921..31056,31167..31254,  
32763..32882,33620..33699,34090..34179,43449..43564,  
43788..44261)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-002"  
join(3317..3450,4893..4974,7738..7790,9656..9844,  
14861..14988,28384..28470,30921..31056,31167..31254,  
32763..32882,33620..33699,34090..34179,43449..43564,  
43788..44261)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-002"  
/product="protein kinase C, zeta"  
/note="match: CDNAS: Em:AK097627.1"  
14579..14721  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-009"  
/note="single clone region. Assembly confirmed by  
restriction digest data."  
join(34127..34179,40698..40780,43449..43564,43788..43982)  
/gene="PRKC2"  
/locus\_tag="RP11-181G12.1-012"  
join(34127..34179,40698..40780,43449..43564,43788..43982)  
mRNA

Query Match 10.6%; Score 212.8; DB 8; Length 179836;  
Best Local Similarity 79.2%; Pred. No. 5.2e-38;  
Matches 290; Conservative 0; Mismatches 72; Indels 4; Gaps 3;

gene  
1156 AAGAGGCGCATTCAGTGTGAAATGTTTCAGAGAGATTGATTGATTGTTTCTCAGA 1215  
DB 46107 AAAAGAGGTGTTCCAAAGAGGTTTTCGAAGCCATTTCCTTTTTTTTCTTT 46166  
QY 1216 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTCACCCAGCGCTGAGTACA 1275  
DB 46167 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTCACCCAGCGCTGAGTACA 46226  
QY 1276 GTGCGTGTGCTCGGCTCAGCGCACTCTGCTCCGAGTTTCAGGCACTCTCTG-- 1333  
DB 46227 GTGCG-GAATCTCGGCTCAGCGCACTCTCCTCCGGGTTTAAGCAATTCCTCTGCC 46285  
QY 1334 TCAGCTTCCCGAATAGCTGGATTACAGCGCATGACAC-CATGCTTAATTTTGTAT 1392

DB 46286 TCACCTCCCGCAGTAGCTGGATTACAGGCACCTGCCCAAGCCCGCTAATTTTGTAT 46345  
QY 1393 TTTTAGTAGAGACAGAGTTTGCCATGTTGACCGAGCTTGCTTGAATCTTGACTTCAG 1452  
DB 46346 TTTTAGTAGAGACAGCGGTTTACCATGTTTCCAGGCTGAGCTTGAATCTTGACTTCAG 46405  
QY 1453 GTGATCCACCCAGCTCAGCCCTCCCAAGCACTGGAGTTTACAGGATGACCACTGCGCC 1512  
DB 46406 GTGATCCACCTGCGCTCAGCCCTCCCAAGTTCTGGGATTACAGCGGTGACTACCGGCC 46465  
QY 1513 AGCTTG 1518  
DB 46466 GGCTTG 46471

RESULT 30  
AC068198  
LOCUS  
DEFINITION  
Homo sapiens chromosome 1 clone RP11-290B2, WORKING DRAFT SEQUENCE,  
10 unordered pieces.  
ACCESSION  
AC068198  
VERSION  
AC068198.11 GI:9910748  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULFILLTOP.  
SOURCE  
Homo sapiens  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo  
1 (bases 1 to 213432)  
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,  
Federispiet,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,  
Mao,J., Lam,B., Marthe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,  
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelm,J.,  
Yu,S. and Davis,R.W.  
Unpublished  
2 (bases 1 to 213432)  
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,  
Federispiet,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,  
Mao,J., Marthe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,  
Ramirez,D., Wilhelm,J., Yu,S. and Davis,R.W.  
Direct Submission  
Submitted (30-Apr-2000) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
On Aug 27, 2000 this sequence version replaced gi:8567772.  
----- Genome Center  
Center: Stanford DNA Sequencing and Technology Development  
Center  
Center code: SDSTDC  
Web site: http://sequence-www.stanford.edu/group/human/  
Contact: hum-info@sequence.stanford.edu  
----- Project Information  
Center project name: 887  
Center clone name: RP11-290B2  
----- Summary Statistics  
Sequencing Vector: M13mp18; X02513  
Chemistry: Dye-primers; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.99019  
Consensus quality: 205371 bases at least Q40  
Consensus quality: 208312 bases at least Q30  
Consensus quality: 209125 bases at least Q20  
Insert size: 197597; agarose-fp  
Insert size: 212532; sum-of-contigs  
Quality coverage: 8.3x in Q20 bases; agarose-fp  
Quality coverage: 7.7x in Q20 bases; sum-of-contigs.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence







```

54791. .55033)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-004"
join(25443. .25513,29191. .29375,50504. .50551,51878. .51987,
54791. .55033)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-004"
/ product="novel protein"
/ note="match: ESTs: Em:BU935571.1"
join(25445. .25513,50504. .50551,51878. .51987,53796. .54036,
54791. .54991)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-005"
/ locus_tag="RP5-1056L3.2-005"
/ product="novel protein"
/ note="match: ESTs: Em:CB157536.1 Em:CB162365.1
Em:CB163300.1"
join(25450. .25513,50504. .50551,51878. .51987,54791. .54805)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-001"
/ standard_name="OTTHUMP00000002794"
/ note="match: proteins: Tr:AAH5791 Tr:Q96G68"
/ codon_start=1
/ product="novel protein"
/ protein_id="CA11917.1"
/ db_xref="GI:56203284"
/ db_xref="InterPro:IPR007512"
/ translation="MSBSLGRKMDRCIADAVKIGTGFGLGIVSLTFPRRMPLA
FGSGMGIAVGNCOHDFQAPYLHGXVKEQEQ"
join(25456. .25513,29191. .29708)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-006"
/ join(25456. .25513,29191. .29708)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-006"
/ product="novel protein"
/ note="match: ESTs: Em:AL602952.1"
join(25474. .25513,26238. .26257,50504. .50551,51878. .51987,
54791. .55010)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-007"
join(25474. .25513,26238. .26257,50504. .50551,51878. .51987,
54791. .55010)
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-007"
/ product="novel protein"
/ note="match: ESTs: Em:BM504924.1"
32581. .32758
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-003"
/ note="sequence confirmed by AP003160 sequenced by RIKEN"
36518. .36972
/ locus_tag="RP5-1056L3.3-001"
/ pseudo
/ codon_start=1
/ product="ribosomal protein S14 (RPS14) pseudogene"
42637. .42645
/ gene="RP5-1056L3.2"
/ locus_tag="RP5-1056L3.2-003"
/ note="1329 bases of Tn10 transposon (J01829) removed
here. This sequence represents the duplicated flanking
sequence of the Tn10."
55011. .55016
/ gene="RP5-1056L3.2"

```

Query Match	10.6%	Score 212;	DB 8;	Length 163848;
Best Local Similarity	78.5%;	Pred. No. 8e-38;		
Matches 292;	Conservative 0;	Mismatches 75;	Indels 5;	Gaps 3;
polYA_site	/locus_tag="RP5-1056L3.2-004"	55027	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-004"	56113	/gene="RP5-1056L3.2"	
polYA_signal	/locus_tag="RP5-1056L3.2-001"	58198	/gene="RP5-1056L3.2"	
polYA_site	/locus_tag="RP5-1056L3.2-001"	58224	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	62566	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	65798	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70460	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70645	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70646	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70646	/gene="RP5-1056L3.2"	
Query Match	10.6% <td>Score 212; <td>DB 8; <td>Length 163848; </td></td></td>	Score 212; <td>DB 8; <td>Length 163848; </td></td>	DB 8; <td>Length 163848; </td>	Length 163848;
Best Local Similarity	78.5% <td>Pred. No. 8e-38; <td> <td> </td></td></td>	Pred. No. 8e-38; <td> <td> </td></td>	<td> </td>	
Matches 292; <td>Conservative 0; <td>Mismatches 75; <td>Indels 5; <td>Gaps 3; </td></td></td></td>	Conservative 0; <td>Mismatches 75; <td>Indels 5; <td>Gaps 3; </td></td></td>	Mismatches 75; <td>Indels 5; <td>Gaps 3; </td></td>	Indels 5; <td>Gaps 3; </td>	Gaps 3;
polYA_site	/locus_tag="RP5-1056L3.2-004"	55027	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-004"	56113	/gene="RP5-1056L3.2"	
polYA_signal	/locus_tag="RP5-1056L3.2-001"	58198	/gene="RP5-1056L3.2"	
polYA_site	/locus_tag="RP5-1056L3.2-001"	58224	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	62566	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	65798	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70460	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70645	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70646	/gene="RP5-1056L3.2"	
misc_feature	/locus_tag="RP5-1056L3.2-001"	70646	/gene="RP5-1056L3.2"	

Bouck, J., Bowie, S., Briteva, M., Brown, E., Brown, M., Bryant, N. P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.P., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, F., Ferraguto, D., Flagg, N., Ford, V., Foster, P., Frantz, P.,  
Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J., Gu, J., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Holloway, C.,  
Hollins, B., Homs, F., Howard, S., Huber, T., Hulik, S., Hume, J.,  
Jackson, L.B., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
Kovar, C., Krivoy, C., Kuresh, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichteage, O., Lieu, C., Liu, J., Liu, W.,  
Loulsgaard, R., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapu, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M.,  
Mei, G., Metzger, M., Miner, G., Morris, Z., Mitchell, T., Monbaba, K.,  
Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D.,  
Nelsson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N.,  
Nikerson, E., Nwokkenko, S., Ogih, M., Okunou, G., Oragunye, N.,  
Oviedo, R., Pace, A., Payton, B., Peary, J., Perez, L., Peters, L.,  
Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M.,  
Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S.,  
Scott, G., Shen, H., Shoshitelli, N., Sisson, I., Sodexyren, E.,  
Sonaite, T., Sparks, A., Stanley, L., Stone, H., Sutton, A., Svatek, A.,  
Tabor, P., Tamerisa, A., Tameris, K., Tang, H., Tansey, J., Taylor, C.,  
Taylor, T., Telford, B., Thomas, N., Thomas, S., Uman, K., Vasquez, L.,  
Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S.,  
Warren, R., Washington, C., Wallington, S., Williams, G.,  
Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.,  
Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G. and  
Gibbs, R.

**TITLE** Direct Submission  
**REFERENCE** Unpublished  
**AUTHORS** 2 (bases 1 to 70446)  
**TITLE** Worley, K.C.  
**JOURNAL** Direct Submission  
Submitted (18-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 70446)  
**REFERENCE** Worley, K.C.  
**AUTHORS** Direct Submission  
**TITLE** Submitted (22-FEB-2002) Human Genome Sequencing Center, Department  
**JOURNAL** of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 70446)  
**REFERENCE** Worley, K.C.  
**AUTHORS** Direct Submission  
**TITLE** Submitted (23-MAR-2002) Human Genome Sequencing Center, Department  
**JOURNAL** of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 70446)  
**REFERENCE** Worley, K.C.  
**AUTHORS** Direct Submission  
**TITLE** Submitted (09-APR-2002) Human Genome Sequencing Center, Department  
**JOURNAL** of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Mar 23, 2002 this sequence version replaced gi:18449819.  
**COMMENT** INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu

```

Features listing.

ANNORATION OF FEATURES:
STSs are identified using ePCR (Genome Ref. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

FEATURES
source
QUALSTAT-REPORT.
location/Qualifiers
1..70446
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-367K14"
complement(1..2004)
/note="overlaps bases 1..2004 of clone AC072044"
/function="clone overlap"
246..408
/rpt_family="LTR33A"
1054..1347
/rpt_family="AluJo"
/rpt_family="AluJo"
1348..1445
/rpt_family="(TACA)n"
2650..2684
/rpt_family="AT_rich"
complement(3610..3835)
/rpt_family="WER46C"
4100..4339
/rpt_family="WER46A"
complement(4362..4519)
/rpt_family="MIR"
5115..5240
/standard_name="73561"
6259..6338
/rpt_family="AluJ/PLAM"
8023..8086
/rpt_family="AT_rich"
complement(9156..9359)
/rpt_family="WER44D"
complement(9346..9459)
/rpt_family="WER44D"
11171..11222
/rpt_family="(CA)RAT)n"
complement(16175..16226)
/rpt_family="MIR"
complement(16218..16416)
/rpt_family="L2"
16424..19352
/rpt_family="11M4"
complement(19379..19472)

```

```
repeat_region /rpt_family="AluSc"
19473..19833
repeat_region /rpt_family="L1MB8"
complement(19887..20219)
repeat_region /rpt_family="L2"
complement(20237..20287)
repeat_region /rpt_family="L2"
20288..20321
repeat_region /rpt_family="(TTTTTA)n"
complement(20322..20678)
repeat_region /rpt_family="L2"
20679..20951
repeat_region /rpt_family="Aluub"
20952..20977
repeat_region /rpt_family="(TAAA)n"
complement(20978..21365)
repeat_region /rpt_family="L2"
21930..21956
repeat_region /rpt_family="AT_rich"
23899..24161
repeat_region /rpt_family="AluSg"
25101..25233
repeat_region /rpt_family="MIR"
25489..25513
repeat_region /rpt_family="AT_rich"
27792..27816
repeat_region /rpt_family="(T)n"
29516..29542
repeat_region /rpt_family="(TA)n"
30921..31046
repeat_region /rpt_family="MER30"
31047..31360
repeat_region /rpt_family="AluY"
31361..31461
repeat_region /rpt_family="MER30"
33737..33901
repeat_region /rpt_family="MIR"
35270..35310
repeat_region /rpt_family="(TGAA)n"

Query Match 10.6%; Score 211.8; DB 8; Length 70446;
Best Local Similarity 66.5%; Pred. No. 1e-37;
Matches 364; Conservative 0; Mismatches 177; Indels 6; Gaps 4;

1203 ATGTTTTCAGATCTTTTATTTTATTTTGAACAAGCTCACTTGTGACCC 1262
62146 AATAATCTCTCTTTTATTTTATTTTATTTTGAAGCAGAGTCTGCTGTGCCC 62205
1263 AGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTGCTCCAGTTCAAG 1322
62206 AAGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTGCTCCAGTTCAAG 62264
1323 CGATTTCTCCTG--TCAGCTTCCGAAATAGCTGGATTACAGGCGCATGACACAC--ATG 1378
62265 CGATTTCTCCTGCTTACCTCCGAGTAGCTGGATTACAGGCGCATGACACACCTTCCG 62324
1379 CCTATTTTGTATTTTATTTAGTACAGAGTGGCCATGTTGACACAGCTTGCCTTGA 1438
62325 GCTAATTTTGTATTTTATTTAGTACAGAGGTTTACCAATGTTGCGCAGGCTGCTTCA 62384
1439 ACTCTGACTTCAAGTATCCACCCAGCTCAGCTCCCAAGACATGGGATTACAGGCAAT 1498
62385 ACTCTGACTTCAAGTATCCACCACTCAGCTCCCAAGACATGGGATTACAGGCAAT 62444
1499 GAGCCACCGTCCAGGCTGTTTCTCAGATCTGTATTTGT-TTCTGAAGCCTTCATT 1557
62445 AAGCCACCGTCCAGGCTGTTTCTCAGATCTGTATTTGT-TTCTGAAGCCTTCATT 62504
1558 CATCTCTTATTTATTTTGAAGTATACACTAGTAAAGTTTAAACATCAAAATAT 1617
62505 AATAAGACCTCCCTTCTCTGAAAAACGTGATTTTATGAAAAAGTTATTTAAACCTTAAG 62564
1618 CTTTGGAAAAATTCCTGTTCTTCTTATTTCTACAAAAATATATGTTCAGTATAGCTGAT 1677
```

```
Db 62565 TACCAGATGAGCTTAAAAATATGATATATTTTAAATTTCTATGACATCTAGGAAGAT 62624
Qy 1678 GTTATGTTTCTTCAATATATTCATTTCTCTCTCAGAAATTAATCTAGCTATTTCT 1737
Db 62625 AATATGATATACACATTTATACACATATATTTACTCATTTATTCATATATATTTT 62684
Qy 1738 TATTGAA 1744
Db 62685 TTTAGAA 62691

RESULT 34
AC117477
LOCUS
DEFINITION Homo sapiens 3 BAC RP11-631B21 (Roswell Park Cancer Institute Human
AC117477
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 118135)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbata,J., Benton,J., Binage,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Francis,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.B., Jacobson,B., Jia,Y., Johnson,R., Joliver,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryak,T.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Lousaged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,B., Massey,B., Mawhney,B., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neel,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nockenkw,S., Ogunu,M., Okunnu,G., Oragunye,N.,
Oyedeo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,B., Pu,L.L., Quiles,M., Ren,Y., Rives,S.,
Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S.,
Scott,G., Shen,H., Shoshchikov,N., Sisson,I., Sodergren,B.,
Soniak,T., Sparks,A., Stanley,H., Stone,H., Sutton,J., Taylor,C.,
Taylor,T., Telifrod,B., Thomas,N., Thomas,S., Umami,K., Vasquez,L.,
Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,K., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorilla,S., Naylor,S.L., Weinstein,G. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 118135)
Worley,K.C.
Submitted (10-APR-2002) Human Genome Sequencing Center, Department
```

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 118135)  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (24-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 118135)  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (27-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 118135)  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (01-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 1, 2002 this sequence version replaced gi:20334525.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

## ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

FEATURES  
source  
Location/Qualifiers  
1. .118135  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-631B21"  
9. .106  
/rpt\_family="(TAGA)n"  
147. .153  
/note="1331bp IS2 transposon:ECGALLI"  
/function="bacterial transposon excised"  
1311. .1345  
repeat\_region  
/rpt\_family="AT\_rich"

repeat\_region complement (2271. .2496)  
/rpt\_family="MER46C"  
repeat\_region 2761. .3000  
/rpt\_family="MER46A"  
repeat\_region complement (3023. .3180)  
/rpt\_family="MIR"  
STS 3776. .3901  
/standard\_name="73561"  
repeat\_region 4920. .4939  
/rpt\_family="Alu/FLAM"  
6684. .6747  
/rpt\_family="AT\_rich"  
repeat\_region complement (7817. .8020)  
/rpt\_family="MER44D"  
repeat\_region complement (8007. .8120)  
/rpt\_family="MER44D"  
9832. .9883  
/rpt\_family="(CATATA)n"  
repeat\_region complement (14836. .14887)  
/rpt\_family="MIR"  
repeat\_region complement (14879. .15077)  
/rpt\_family="L2"  
15085. .18013  
/rpt\_family="L1M4"  
repeat\_region complement (18040. .18133)  
/rpt\_family="AluSc"  
18134. .18494  
/rpt\_family="L1MB8"  
repeat\_region complement (18548. .18880)  
/rpt\_family="L2"  
complement (18898. .18948)  
/rpt\_family="L2"  
18949. .18982  
/rpt\_family="(TTTAA)n"  
repeat\_region complement (18983. .19339)  
/rpt\_family="L2"  
19340. .19612  
/rpt\_family="AluDb"  
19613. .19638  
/rpt\_family="(TAA)n"  
repeat\_region complement (19639. .20026)  
/rpt\_family="L2"  
20591. .20617  
/rpt\_family="AT\_rich"  
repeat\_region 22560. .22822  
/rpt\_family="AluSg"  
23762. .23894  
/rpt\_family="MIR"  
24150. .24174  
/rpt\_family="AT\_rich"  
26453. .26477  
/rpt\_family="(T)n"  
28177. .28203  
/rpt\_family="(TA)n"  
29582. .29707  
/rpt\_family="MER30"  
repeat\_region 29708. .30021  
/rpt\_family="AluY"  
30022. .30122  
/rpt\_family="MER30"  
32398. .32562  
/rpt\_family="MIR"  
33931. .33971  
/rpt\_family="(TGA)n"  
repeat\_region complement (34508. .34565)  
/rpt\_family="MIR"  
35991. .36054  
/rpt\_family="GA-rich"

Query Match 10.6%; Score 211.8; DB 8; Length 118135;  
Best Local Similarity 66.5%; Pred. No. 9.3e-38;  
Matches 364; Conservative 0; Mismatches 177; Indels 6; Gaps 4;

[illegible]

RESULT 35				
AC025415/C				
LOCUS	AC025415	181505 bp	DNA	linear
DEFINITION	Homo sapiens chromosome 3 clone RP11-367K14 map 3,			HTG 24-AUG-2002
ACCESSION	SEQUENCE, 23 unordered pieces.			WORKING DRAFT
VERSION	AC025415			
SOURCE	AC025415.3 GI:7770520			
ORGANISM	HTG; HTGS_PHASE1; HTGS_DRAFT.			
	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
	Homniidae; Homo.			
REFERENCE	1 (bases 1 to 181505)			
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.			
TITLE	Homo sapiens chromosome 3, clone RP11-367K14			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 181505)			
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F., Boguski,M.V., Boulton,A., Boulikas,G., Brown,A., Burtick,G., Campilongo,A., Castle,A., Chappel,Y., Colangelo,M., Collins,S., Collins,R.A., Cooke,P., DeRubeis,J., Dewar,K., Diaz,J.S., Dodgson,S., Donnelly,D., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardy,J., Ginde,S., Glendon,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Haeussler,B., Heaford,A., Horton,L., Howland,T.C., Iliev,I., Johnson,S., Jones,C., Kann,L., Karst,A., Klein,J., Laboucque,K., Lamazares,R., Landers,T., Lechczky,J., Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Margulies,N., McCarthy,M., McKean,P., McGuire,K., McKernan,K., McPherson,R., McPheron,J., Meneses,L., Minova,T., Miranda,C., Mlenaga,V., Morrow,J., Mudryk,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,			

TITLE	JOURNAL	REFERENCE
TITLE	JOURNAL	AUTHORS
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stenge-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J., Tirrell, A., Traverser, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	Direct Submission	
Submitted (08-MAR-2000)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	3 (bases 1 to 181505)
Anderson, S., Baldwin, J., Barn, N., Bastien, V., Bedd, F., Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N., Boguski, V. L., Bouhgalter, B., Brown, A., Burckett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Donlu, M., Doyle, M., Ferreira, P., Fitzhugh, N., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marguis, N., McCarthy, M., McEwan, P., McGuir, A., McKernan, K., McHeeters, R., Melchior, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stenge-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J., Tirrell, A., Traverser, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	Direct Submission	
Submitted (24-AUG-2002)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
On May 12, 2000 this sequence version replaced g1:1528375.		

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L8179
Center clone name: 367 K 14
-----
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166432 bases at least Q40
Consensus quality: 174531 bases at least Q30
Consensus quality: 177652 bases at least Q20
Insert size: 185000; agarose-fp
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1123: contig of 1123 bp in length
*
*      1124      1223: gap of 100 bp
*
*      1224      2181: contig of 958 bp in length
*
*      2182      2281: gap of 100 bp
*
*      2282      4712: contig of 2431 bp in length
*
*      4713      4812: gap of 100 bp

```

```
* 4813 8198: contig of 3386 bp in length
* 8198: gap of 100 bp
* 8298: contig of 3111 bp in length
* 8299 11409: gap of 100 bp
* 11410 11509: contig of 3667 bp in length
* 11510 15116: contig of 3667 bp in length
* 15117 15216: gap of 100 bp
* 15217 19171: contig of 3955 bp in length
* 19172 22792: contig of 3521 bp in length
* 22793 22892: gap of 100 bp
* 22893 27631: contig of 4739 bp in length
* 27632 27731: gap of 100 bp
* 27732 33828: contig of 6097 bp in length
* 33829 40616: contig of 6688 bp in length
* 40617 40716: gap of 100 bp
* 40717 45477: contig of 4761 bp in length
* 45478 45577: gap of 100 bp
* 45578 53703: contig of 8126 bp in length
* 53704 53803: gap of 100 bp
* 53804 61654: contig of 7851 bp in length
* 61655 72106: gap of 100 bp
* 72107 72206: contig of 10352 bp in length
* 72207 81246: gap of 100 bp
* 81247 81346: gap of 100 bp
* 81347 89996: contig of 8650 bp in length
* 89997 90096: gap of 100 bp
* 90097 100217: contig of 10121 bp in length
* 100218 100317: gap of 100 bp
* 100318 114540: contig of 14223 bp in length
* 114541 114640: gap of 100 bp
* 114641 129350: contig of 14710 bp in length
* 129351 129450: gap of 100 bp
* 129451 143367: contig of 13917 bp in length
* 143368 143467: gap of 100 bp
* 143468 161284: contig of 17817 bp in length
* 161285 161384: gap of 100 bp
* 161385 181505: contig of 20121 bp in length.
```

## FEATURES

source

```
1. 181505
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-367K14"
/clone_1fb="RP11 Human Male BAC"
1. 1123
/note="assembly_fragment"
1124. 1223
/estimated_length=100
1224. 2181
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
2182. 3281
/estimated_length=100
2282. 4712
/note="assembly_fragment"
4713. 4812
/estimated_length=100
4813. 8198
/note="assembly_fragment"
8199. 8298
/estimated_length=100
8299. 11409
/note="assembly_fragment"
11410. 11509
/estimated_length=100
11510. 15116
/note="assembly_fragment"
15117. 15216
```

Query Match 10.6%; Score 211.8; DB 14; Length 181505;  
Best Local Similarity 66.5%; Pred. No. 8.8e-38;  
Matches 364; Conservative 0; Mismatches 177; Indels 6; Gaps 4;

```
misc_feature /estimated_length=100
15217..19171
/note="assembly_fragment"
gap 19172..19271
/estimated_length=100
misc_feature 19272..22792
/note="assembly_fragment"
gap 22793..22892
/estimated_length=100
misc_feature 22893..27631
/note="assembly_fragment"
gap 27632..27731
/estimated_length=100
misc_feature 27732..33828
/note="assembly_fragment"
gap 33829..33928
/estimated_length=100
misc_feature 33929..40616
/note="assembly_fragment"
gap 40617..40716
/estimated_length=100
misc_feature 40717..45477
/note="assembly_fragment"
clone_end:17
```

RESULT 36  
AC161624 190264 bp DNA linear HTG 17-MAY-2005  
LOCUS AC161624



DEFINITION	REFERENCE
Pan troglodytes chromosome UNKNOWN clone CH251-149E7, ***	
SEQUENCING IN PROGRESS ***, 32 unordered pieces.	
ACCESSION	AC161624
VERSION	AC161624.1
KEYWORDS	HTG, HTGS, PHASE1.
SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Pan troglodytes
	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
	Hominidae; Pan.
	1 (bases 1 to 190264)
AUTHORS	Wilson,R.K.
TITLE	The sequence of Pan troglodytes clone
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 190264)
TITLE	Wilson,R.K.
JOURNAL	Direct Submission
	Submitted (17-MAY-2005) Genetics, Genome Sequencing Center, 4444
	Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	
	----- Genome Center -----
	Center: Washington University Genome Sequencing Center
	Center code: WUGSC
	Web site:http://genome.wustl.edu
	----- Project Information -----
	Center project name: C_AB0149E07
	----- Summary Statistics -----
	Sequencing vector: M13, 0%
	Sequencing vector: plasmid, 100%
	Chemistry: Dye-terminator Big Dye, 100% of reads
	Chemistry: Dye-terminator Big Dye, 100% of reads
	Assembly program: Phrap; version 0.990319
	Consensus quality: 169768 bases at least Q40
	Consensus quality: 175204 bases at least Q30
	Consensus quality: 178674 bases at least Q20
	-----
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 32 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	1 1180: contig of 1180 bp in length
	1181 1280: gap of unknown length
	1281 2540: contig of 1260 bp in length
	2541 2640: gap of unknown length
	2641 3939: contig of 1299 bp in length
	3940 4039: gap of unknown length
	4040 5503: contig of 1464 bp in length
	5504 5603: gap of unknown length
	5604 7410: contig of 1807 bp in length
	7411 7511: gap of unknown length
	7511 9136: contig of 1626 bp in length
	9137 9237: gap of unknown length
	9237 10445: contig of 1209 bp in length
	10446 10545: gap of unknown length
	10546 12371: contig of 1826 bp in length
	12372 12471: gap of unknown length
	12471 14015: contig of 1544 bp in length
	14016 14115: gap of unknown length
	14116 15377: contig of 1262 bp in length
	15378 15477: gap of unknown length
	15478 17676: contig of 2199 bp in length
	17677 20317: gap of unknown length
	20317 20417: contig of 2541 bp in length
	20418 22581: gap of unknown length
	22581 22681: contig of 2164 bp in length
	22682 26214: gap of unknown length
	26214 26215: contig of 3533 bp in length
	26215 26215: gap of unknown length

	*	26315	29199	contig of 2885 bp in length
	*	29200	32299	gap of unknown length
	*	29200	32273	contig of 2374 bp in length
	*	32274	32373	gap of unknown length
	*	32374	37340	contig of 4967 bp in length
	*	37341	37440	gap of unknown length
	*	37441	41969	contig of 4529 bp in length
	*	41970	42069	gap of unknown length
	*	42070	48594	contig of 6525 bp in length
	*	48595	48694	gap of unknown length
	*	48695	55037	contig of 6343 bp in length
	*	55038	55137	gap of unknown length
	*	55138	60618	contig of 5481 bp in length
	*	60619	60718	gap of unknown length
	*	60719	65847	contig of 5129 bp in length
	*	65848	65947	gap of unknown length
	*	65948	71772	contig of 5825 bp in length
	*	71773	71872	gap of unknown length
	*	71873	79030	contig of 7158 bp in length
	*	79031	79130	gap of unknown length
	*	79131	86151	contig of 7021 bp in length
	*	86152	86251	gap of unknown length
	*	86252	97553	contig of 11302 bp in length
	*	97554	97653	gap of unknown length
	*	97654	106778	contig of 9125 bp in length
	*	106779	106878	gap of unknown length
	*	106879	115928	contig of 9050 bp in length
	*	115929	116028	gap of unknown length
	*	116029	124164	contig of 8136 bp in length
	*	124165	124264	gap of unknown length
	*	124265	138817	contig of 14553 bp in length
	*	138818	138917	gap of unknown length
	*	138918	154701	contig of 15784 bp in length
	*	154702	154801	gap of unknown length
	*	154802	190264	contig of 35463 bp in length
FEATURES				
source				Location/Qualifiers
			1..190264	
			/organism="Pan troglodytes"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:9598"	
			/chromosome="UNKNOWN"	
			/clone="CH251-149B7"	
			1..1180	
misc_feature			/note="assembly_name:Contig20"	
gap			1181..1280	
			/estimated_length=unknown	
misc_feature			1281..2540	
			/note="assembly_name:Contig23"	
gap			2541..2640	
			/estimated_length=unknown	
misc_feature			2641..3935	
			/note="assembly_name:Contig24"	
gap			3940..4039	
			/estimated_length=unknown	
misc_feature			4040..5503	
			/note="assembly_name:Contig25"	
gap			5504..5603	
			/estimated_length=unknown	
misc_feature			5604..7410	
			/note="assembly_name:Contig26"	
gap			7411..7510	
			/estimated_length=unknown	
misc_feature			7511..9136	
			/note="assembly_name:Contig27"	
gap			9137..9236	
			/estimated_length=unknown	
misc_feature			9237..10445	
			/note="assembly_name:Contig28"	
gap			10446..10545	
			/estimated_length=unknown	
misc_feature			10546..12371	
			/note="assembly_name:Contig29"	
gap			12372..12471	

```
misc_feature /estimated_length=unknown
12472..14015
/contig30"
gap 14016..14115
/estimated_length=unknown
misc_feature 14116..15377
/contig31"
gap 15378..15477
/estimated_length=unknown
misc_feature 15478..17576
/contig32"
gap 17577..17776
/estimated_length=unknown
misc_feature 17777..20317
/contig33"
gap 20318..20417
/estimated_length=unknown
misc_feature 20418..22581
/contig34"
gap 22582..22681
/estimated_length=unknown
misc_feature 22682..26214
/contig35"
gap 26215..26314
/estimated_length=unknown
misc_feature 26315..29199
/contig36"
gap 29200..29299
/estimated_length=unknown
misc_feature 29300..32273
/contig37"
gap 32274..32373
/estimated_length=unknown
misc_feature 32374..37340
/contig38"
gap 37341..37440
/estimated_length=unknown
misc_feature 37441..41969
/contig39"
gap 41970..42069
/estimated_length=unknown
misc_feature 42070..48594
/contig40"
gap 48595..48694
/estimated_length=unknown
misc_feature 48695..55037
/contig41"
gap 55038..55137
/estimated_length=unknown
misc_feature 55138..60618
/contig42"
gap 60619..60718
/estimated_length=unknown

Query Match 10.6%; Score 211.6; DB 14; Length 190264;
Best Local Similarity 83.6%; Pred. No. 9.7e-38;
Matches 276; Conservative 0; Mismatches 49; Indels 5; Gaps 3;

QY 1213 AGATCTTTTATTTATTTTATTTTGAACAAGAGTCTCACTTTGTGACCCAGGCTGGAGT 1272
DB 129021 ACAATATCTCTCTTTTATTTTATTTTGAAGTGAAGTCTCTGCTGTCAACCCAGGCTGGAGT 129080
QY 1273 AAGAGGCTGTGCTCGGCTCACTGCAACCTCTGCTCCAGGCTTCAGGCGATTCTCCT 1332
DB 129081 GAG-GGGACAAATCTGGCTCACTGCAATCTCGCTCCAGGTTCAAGCGATTCTCCT 129139
QY 1333 G--TCAGCTTCCGAATAGCTGGAGTTAAGGCGCATGACCAACCATGCTAATTTTGT 1390
DB 129140 GCCTTAGCTCCCGAGTACTGGGATTAAG-GCATGTGCCACCAACGCTAATTTTGT 129197
QY 1391 ATTTTATGAGACAGAGTTTGCAGTGTGACCAAGGCTTGCCTTGAATCTCTGAATTC 1450
DB 129198 ATTTTATGAGACAGGAGTTTGCAGTGTGACCAAGGCTTGCCTGAATCTCTGAATTC 129257
```

```
QY 1451 AGGTATCCACCACCTGAGCTTCCAAAGCATGGATTAACGATGACCAACCGTGC 1510
DB 129258 AGGTATCCACCACCTGAGCTTCCAAAGCATGGATTAACGATGACCAACCGTGC 129317
QY 1511 CCAGCTGTTTCTCAGATCTCTGATTTGT 1540
DB 129318 CCAGCTGTTTCTCAGATCTCTGATTTGT 129347

RESULT 37
AC008761 226170 bp DNA linear PRI 13-JUL-2002
LOCUS Homo sapiens chromosome 19 clone CTD-3149D2, complete sequence.
DEFINITION AC008761
VERSION AC008761.8 GI:21743747
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 226170)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 226170)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 226170)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 226170)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jul 13, 2002 this sequence version replaced gi:20177645.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
www.shgc.stanford.edu
Finishing Completed at Stanford Human Genome Center
Quality: phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.4.

FEATURES
source 1..226170
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/location="CTD-3149D2"

ORIGIN
Query Match 10.6%; Score 211.6; DB 8; Length 226170;
Best Local Similarity 81.3%; Pred. No. 9.4e-38;
Matches 270; Conservative 0; Mismatches 59; Indels 3; Gaps 2;

QY 1207 TTCTCAGATCTTTTATTTTATTTTGAACAAGAGTCTCACTTTGTGACCCAGGC 1266
DB 144295 TCTCTCTTTTATTTTATTTTGAAGCAGAGTCTGCTGTCAACCCAGGC 184354
QY 1267 TGAAGTACAGTGGCTGTGCTCACTGCAACCTCTGCTCCAGGTTCAAGCGAT 1326
DB 144355 TGAAGTACAGTGG-TCAGTCTCTGCTCACTGCAACCTCTGCTCCAGGTTCAAGCGAT 184413
QY 1327 TCTCCG--TCAGCTTCCGAATAGCTGGAGTTAAGGCGCATGACCAACCATGCTAAT 1384
DB 144414 TCTCTGCTCAGCTCCCAAGTACTGGGATTAAGGCGCTTGCACCATGCTCGGCTT 184473
```

Oy	1385	TTTTGGTATTTTAACTAAGACAGAGTTTCCCATTTGTCACAGGCTTGCCTGAACCTCT	1444
Db	184474	TTTGTATTTTACACAGACAGAGGTTTTCACATGGTGGCCAGGCTGGCTTGAACCTCT	184533
Oy	1445	GACCTCAGGTGATCCACCCACCTCAGCTCCAAAGACCTGGAGTTACAGCATGAGCCA	1504
Db	184534	GACCTCAGGTGATCCGCCCGCTTACGCTCCCAAAGTGTGGAGTTACAGGATGGGCCA	184559
Oy	1505	CCGTGCCAGGCTGTTTCTTCAGATCCTGAT	1536
Db	184594	CCGTGCCAGGCTCCCTCAATTGAGTTCCGCT	184625
RESULT 38			
LOCUS	D88268	44553 bp	DNA linear PRI 26-MAR-2002
DEFINITION	Homo sapiens immunoglobulin lambda gene locus DNA, clone:47H9.		
ACCESSION	D88268		
VERSION	D88268.1	GI:2114305	
KEYWORDS	immunoglobulin light chain.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J.L., Wang, J. and Shimizu, N.		
TITLE	One-megabase sequence analysis of the human immunoglobulin lambda gene locus		
JOURNAL	Genome Res. 7 (3), 250-261 (1997)		
PubMed	9074928		
REFERENCE	2 (bases 1 to 44553)		
AUTHORS	Shimizu, N.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-OCT-1996) Nobuyoshi Shimizu, Keio University, School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@med.med.keio.ac.jp, Tel:03-3351-2370, Fax:03-3351-2370)		
FEATURES	Location/Qualifiers		
source	1..44553		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="22"		
	/map="22q11.2"		
	/clone="47H9"		
	/cell_type="human/rodent somatic cell hybrid"		
	/clone_id="specific cosmid library; LU22NC03"		
	1..44553		
	/note="region V upstream"		
	/db_xref="IMGT/LIGM:D88268"		
ORIGIN			
Query Match	10.5%; Score 211; DB 8; Length 44553;		
Best Local Similarity	72.4%; Prad. No. 1.6e-37;		
Matches 315; Conservative 0; Mismatches 115; Indels 5; Gaps 3;			
Oy	1104	TGAAATGCTTCAGAACCCCAATTATGAGAAATTTTATGACCATGTGTGACAAAGAGAGG	1163
Db	31604	TGTAAAGATATCTCAATCAACAAATGTGTCTTGTTTATCTGATCTAAACAGCAAT	31545
Oy	1164	CCATTCAGATGTTGAATTTGTTTCAGAGAAAGATTTGATATGTTTCTCAGATCTTTTAA	1223
Db	31644	TTTATGATATGATCAAAAGTTTGAGCAGATATTTATATGATATATACAGAAAGATTATTT	31485
Oy	1224	TTTTTATTTTTTTGAAACAGAGTCTCATCTTGTCACCAGGCTGAGTACAGTGCT	1283
Db	31484	TTTTTTTTTTTTTTGAAAGAGCTTGCTCTGTCAAGCAGGCTGGAGTGCAGTGCC-AT	31426
Oy	1284	GGTCCGGCTCACGCAACCTCTGCGCCCAAGTGTAAAGCATCTTCCTG--TCAGCTTC	1341
Db	31425	GATCTCGGCTCACGCAACCTCTGCGCTCTGCGGCTCAAGCAATTTCTTCCTCAGGCTC	31366

QY	1342	CCGAATGACTGGAGTTTACAGGCCCATG-C-ACACATGCTTAATTTTGTATTTTACT	1399
DB	31365	CCGAGTAGCTAAGATTACAGGGGTATATCCACCATGCCCGGCTAATTTTGTATTTTACT	313066
QY	1400	AGAGACAGAGTTTCGCAGTGTGACCAAGCTTGCTTGAATCTCTGACTTCAGGTGATCC	1459
DB	31305	AGAGACGGGGTTTACCATGTTGTGGCAGAGGCTGGTCTTGAATCTCTGACTTCAGGTGATCC	312466
QY	1460	ACCACCTGACCTCCCAAGACCTGGGATTATAGGCAATGAGCAACCGGCCAGCTGT	1519
DB	31245	ACCACCTGACCTCCCAAGAGTCTGGGATTACAGGCGTAGGCAACGTCGCCGCGCAGA	311866
QY	1520	TTTCTGACATCTGT	1534
DB	31185	AGTATTAATTTT	31171
RESULT 39			
AF002893/c			
LOCUS	178000 bp	DNA	linear PRI 30-MAR-2002
DEFINITION	Homo sapiens genomic DNA, chromosome 11q clone:RP11-624G17,		
ACCESSION	AP002893		
VERSION	AP002893		
KEYWORDS	AP002893.3 GI:19879819		
SOURCE	HTG.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.		
REFERENCE	1		
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Tsuchi,T., Watanabe,H. and Sakaki,Y.		
TITLE	Homo sapiens genomic DNA		
JOURNAL	Published Only in Database (2000)		
REFERENCE	2 (bases 1 to 178000)		
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Tsuchi,T., Watanabe,H. and Sakaki,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-OCT-2000) Masahito Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Saitoh-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
COMMENT	On Apr 1, 2002, this sequence version replaced gi:13469698.		
FEATURES	Location/Qualifiers		
SOURCE	1. 178000		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="11"		
	/map="11q"		
	/clone="RP11-624G17"		
ORIGIN			
Query Match	10.5%;	Score 211;	DB 8; Length 178000;
Best Local Similarity	80.5%;	Pred. No. 1.3e-37;	
Matches 284; Conservative	0;	Mismatches 65;	Indels 4; Gaps 3;
QY	1216	TCCTTTTATTTTATTTTGTGAAACAGAGTCTCACTTGTGTCACCGAGCTGAGTACA	1275
DB	18298	TCCTTTTCTCTTTTATTTTGTGAGATGAGTCTGCTCTGTTCGCCAAGGCTGAGTGA	18229
QY	1276	GTGGCTGTGCTTCGGCTACTGCAACCTCTGCTCCAGGTTCAAGGATCTCTCG--	1333
DB	18238	GTGG-TACATCTCGGCTTACGCAACCTCTGCTCCAGGTTCAAGATTTCTCTCC	18180
QY	1334	TCAGCTCCCGAATAGCTGGGATTAACAGCGCATGACAC-CATGCTAATTTTGTAT	1392
DB	18179	TCAGCTCCCGAGTAGCTGGGATTAACAGCGCCGCGACACACCTGCGTAAATTTTGTAT	18120
QY	1393	TTTATGAGACAGAGTTTCGCATGTTGACCAAGGCTTGCTTGAATCTCTGACTTCAG	1452

|||||  
Db 18119 TTTTAGTAGAGACAGGGTTTGGCCATGTTGGCCAGGCTGTCTAGAACTCCCAACCTCAG 18060  
Qy 1453 GTGATCCACCCACCTCAGCCTCCCAAGACACTGGATTACAGGCATGAGCCACCGTGCC 1512  
Db 18059 GTGATCCACCCACCTCAGCCTCCCAAGACTGGATTACAGGCCTGTGAGACACCGTGCC 18000  
Qy 1513 AGCGTGTCTCAGATCCTGTAATTTGTTTCTGAAGCCTTCATTTCTATCTTC 1565  
Db 17999 AGCCAGAGGCTTATTTCTTGATTAATTAATCCAGCTCAGGATTTCTTAC 17947  
RESULT 40  
CR589878/c 87071 bp DNA linear PRI 14-SEP-2004  
LOCUS Human DNA sequence from clone XX-HCC1954\_10P22, complete sequence.  
ACCESSION CR589878 GI:51966190  
VERSION CR589878.6  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 87071)  
Peñan, S.  
Direct Submission  
Submitted (14-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Sep 9, 2004 this sequence version replaced GI:51950497.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMBPP; Information  
on the WORMBPP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/XX-HCC1954\_10P22  
is from a Cancer cell lines BACS library VECTOR: pBACe3.6\_BanDI.  
FEATURES  
source  
1..87071  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="XX-HCC1954\_10P22"  
/clone\_1ib="Cancer cell lines BACS"  
ORIGIN  
Query Match 10.5%; Score 210.8; DB 8; Length 87071;  
Best Local Similarity 79.9%; Pred. No. 1.6e-37;  
Matches 286; Conservative 0; Mismatches 67; Indels 5; Gaps 3;  
Qy 1188 GAGAACTATTGATATGTTTCTCAGACTTTTATTTATTTTGAAGACAGT 1247  
Db 15944 GACATGAAGTCTCTATATATTTCTATCTTTTATTTTGAAGACAGT 15885

Qy 1248 CTCACCTTGTACCCAGGCTGAGTACAGTGGCTGTGCTCACTGCAACTCTG 1307  
Db 15884 CTCACCTGTACCCAGGCTGAGACAGTGG-TCAGATTTCGCTCACTCAACTCTCG 15826  
Qy 1308 CTTCCAGGCTTAAAGCATTTCTCTG--TCAGCTTCCGAATAGCTGGATTACAGGGCC 1365  
Db 15825 CTTCCAGGCTTAAAGCATTTCTCTGCTCAGCTCCCGAGTGGCTGGGATTACAGGTGC 15766  
Qy 1366 ATGACACCATGTC--CTAATTTTGTATTTTATGATAGACAGAGTTTGGCATGTGA 1423  
Db 15765 AGCCACACATCCAGGTAATTTTGTATTTTATGTTAGACAGGTTTACCATGTTGG 15706  
Qy 1424 CCAGGCTTGCCTTGAATCTGACTTCAAGTATCCACCACTCAGCCTCCCAAGAC 1483  
Db 15705 CCAAGCTGTGTTTGAATCTGACTTCAAGTATCCACCACTTGGCTCCCAAGTGC 15646  
Qy 1484 TGGATTACAGGATGAGCCACCGTGCCAGCTGTTTCTCAGATCCTGTAATTTGT 1541  
Db 15645 TGGATTACAGGATGAGCCACCAAGCTGCGCTAATTTCTACTTTTAAATGTTT 15588  
RESULT 41  
CR954268/c 107537 bp DNA linear PRI 03-MAY-2005  
LOCUS Human DNA sequence from clone XX-HCC1954\_40B13, complete sequence.  
ACCESSION CR954268  
VERSION CR954268.1 GI:63025095  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 107537)  
Hunter, G.  
Direct Submission  
Submitted (03-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMBPP; Information  
on the WORMBPP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/XX-HCC1954\_40B13  
is from a Cancer cell lines BACS library VECTOR: pBACe3.6\_BanDI.  
FEATURES  
source  
1..107537  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="XX-HCC1954\_40B13"  
/clone\_1ib="Cancer cell lines BACS"

## ORIGIN

Query Match 10.5%; Score 210.8; DB 8; Length 107537;  
Best Local Similarity 79.9%; Pred. No. 1.6e-37;  
Matches 286; Conservative 0; Mismatches 67; Indels 5; Gaps 3;

QY 1188 GAGAAATATTTGATTTGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGT 1247  
DB 75233 GACATGAATTCCTTATTTATTTTCTATATCTTTTATTTTATTTTGAACAGAGT 75174  
QY 1248 CTCACCTTGTCCACGAGCTGAGTACAGTGTCTGCTGCTGCTGCTGCTGCTG 1307  
DB 75173 CTCACCTTGTCCACGAGCTGAGTACAGTGTCTGCTGCTGCTGCTGCTGCTG 75115  
QY 1308 CTTCCAGCTTGAAGGATTTCTCTG--TCAGCTTCCGAAATGCTGGATTAACAGCCG 1365  
DB 75114 CTTCCAGCTTGAAGGATTTCTCTG--TCAGCTTCCGAAATGCTGGATTAACAGCCG 75055  
QY 1366 ATGCACACCATGCG--CTAATTTTGTATTTTATTTAGTACAGATTTGCGCATGTA 1423  
DB 75054 ATGCACACCATGCGATCTGATTTTATTTTATTTTATTTTATTTTATTTTATTTT 74995  
QY 1424 CCAGGCTTGTCTTGAATCTCTGATCTTCAAGTATCCACCTGCTGCTGCTGCTG 1483  
DB 74994 CCAGGCTTGTCTTGAATCTCTGATCTTCAAGTATCCACCTGCTGCTGCTGCTG 74935  
QY 1484 TGGGATTAACAGGATGAGCCACCGTCCAGCTGTTTCTCAGATCTCTGATTTGTT 1541  
DB 74934 TGGGATTAACAGGATGAGCCACCGTCCAGCTGTTTCTCAGATCTCTGATTTGTT 74877

## RESULT 42

LOCUS CR936851 138165 bp DNA linear PRI 26-FEB-2005  
DEFINITION Human DNA sequence from clone XX-HCC1954\_35H09, complete sequence.  
ACCESSION CR936851  
VERSION CR936851.1 GI:60302467  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 138165)  
AUTHORS Kay, M.  
TITLE Direct Submission  
JOURNAL Submitted (25-FEB-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 2 (bases 1 to 138165)  
AUTHORS Kay, M.  
TITLE Direct Submission  
JOURNAL Submitted (25-FEB-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 3 (bases 1 to 138165)  
AUTHORS Kay, M.  
TITLE Direct Submission  
JOURNAL Submitted (25-FEB-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 4 (bases 1 to 138165)  
AUTHORS Kay, M.  
TITLE Direct Submission  
JOURNAL Submitted (25-FEB-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 5 (bases 1 to 138165)  
AUTHORS Kay, M.  
TITLE Direct Submission  
JOURNAL Submitted (25-FEB-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 6 (bases 1 to 138165)  
AUTHORS Kay, M.  
TITLE Direct Submission  
JOURNAL Submitted (25-FEB-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep\_XX-HCC1954\_35H09  
is from a Cancer cell lines BACS library VECTOR: pAC3.6\_BanH1.  
Location/Qualifiers  
1..138165  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="XX-HCC1954\_35H09"  
/clone\_1ib="Cancer cell lines BACS"

## ORIGIN

Query Match 10.5%; Score 210.8; DB 8; Length 138165;  
Best Local Similarity 79.9%; Pred. No. 1.5e-37;  
Matches 286; Conservative 0; Mismatches 67; Indels 5; Gaps 3;

QY 1188 GAGAAATATTTGATTTGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGT 1247  
DB 72260 GACATGAATTCCTTATTTATTTTCTATATCTTTTATTTTATTTTGAACAGAGT 72319  
QY 1248 CTCACCTTGTCCACGAGCTGAGTACAGTGTCTGCTGCTGCTGCTGCTGCTG 1307  
DB 72320 CTCACCTTGTCCACGAGCTGAGTACAGTGTCTGCTGCTGCTGCTGCTGCTG 72378  
QY 1308 CTTCCAGCTTGAAGGATTTCTCTG--TCAGCTTCCGAAATGCTGGATTAACAGCCG 1365  
DB 72379 CTTCCAGCTTGAAGGATTTCTCTG--TCAGCTTCCGAAATGCTGGATTAACAGCCG 72438  
QY 1366 ATGCACACCATGCG--CTAATTTTGTATTTTATTTAGTACAGATTTGCGCATGTA 1423  
DB 72439 ATGCACACCATGCGATCTGATTTTATTTTATTTTATTTTATTTTATTTTATTTT 72498  
QY 1424 CCAGGCTTGTCTTGAATCTCTGATCTTCAAGTATCCACCTGCTGCTGCTGCTG 1483  
DB 72499 CCAGGCTTGTCTTGAATCTCTGATCTTCAAGTATCCACCTGCTGCTGCTGCTG 72558  
QY 1484 TGGGATTAACAGGATGAGCCACCGTCCAGCTGTTTCTCAGATCTCTGATTTGTT 1541  
DB 72559 TGGGATTAACAGGATGAGCCACCGTCCAGCTGTTTCTCAGATCTCTGATTTGTT 72616

## RESULT 43

LOCUS CR938745 164033 bp DNA linear PRI 22-MAR-2005  
DEFINITION Human DNA sequence from clone XX-HCC1954\_32013, complete sequence.  
ACCESSION CR938745  
VERSION CR938745.1 GI:61846497  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 164033)  
AUTHORS Johnson, C.  
TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 2 (bases 1 to 164033)  
AUTHORS Johnson, C.  
TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 3 (bases 1 to 164033)  
AUTHORS Johnson, C.  
TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk

REFERENCE 4 (bases 1 to 164033)  
AUTHORS Johnson, C.  
TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk



TITLE The sequence of Pan troglodytes BAC clone CH251-490K17  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 207664)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (17-FEB-2005) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 207664)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (19-MAR-2005) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 207664)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (13-APR-2005) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 5 (bases 1 to 207664)  
AUTHORS Wilson,R.K.  
TITLE Direct Submission  
JOURNAL Submitted (27-APR-2005) Washington University School of Medicine,  
Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO  
63108, USA  
On Apr 13, 2005 this sequence version replaced gi:60459078.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUSGC  
Web site: http://genome.wustl.edu  
Contact: submissions@watson.wustl.edu  
----- Summary Statistics  
Center project name: C\_AB0490K17

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
http://genome.wustl.edu

## SOURCE INFORMATION:

The CHORI-251 Chimpanzee BAC library has been constructed at the  
Children's Hospital Oakland Research Institute, BACPAC Resources,  
by Dr. Baoji Zhu. DNA was isolated from white blood cells obtained  
from a male chimpanzee (Pan troglodytes, 'Cint', Yerkes #C0471;  
birthdate:6-6-80). The clone and detailed information can be  
obtained from Pieter de Jong and co-workers at  
http://www.bacpac.chori.org.

## FEATURES

## source

This sequence is the entire insert of the clone.

Location/Qualifiers  
1..207664  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="unknown"  
/clone="CH251-490K17"  
/clone\_1ib="CHORI251"  
28714..28766  
/note="Sequence derived from one plasmid subclone."

## ORIGIN

Query Match 10.5%; Score 210.8; DB 8; Length 207664;  
Best local Similarity 82.9%; Pred. No. 1.5e-37;

Matches 277; Conservative 0; Mismatches 52; Indels 5; Gaps 3;  
QY 1197 TTGATTATGTTTCTCAGATCTTTATTTATTTTGAACAGAGTCACTTGG 1256  
|||  
DB 132901 TTTCTTTTGTGTTTCTTTCTTTTATTTATTTTAAACAGAGTCTCCCTTGG 132842  
|||  
QY 1257 TCACCCAGGCTGAGTACAGTGGCTGCTCCGCTCACTCAACTTGCCTCCAGG 1316  
|||  
DB 132841 TCGCCAGGCTGAGTGCAGTGGC-ACGATCTCGGCTACTGCACACTTGCCTCCCGGG 132783  
|||  
QY 1317 TTCAAGCATTCCTCTG--TCAGCTTCCGAATTCCTGGATTACAGGCGCATGC-ACC 1372  
|||  
DB 132782 TTCAAGTATTCCTGCTGCTCCTCAGCTCCTGAGTATTAACAGGCGCCGCATGC 132723  
|||  
QY 1373 ACCATGCTTAATTTTGTATTTTGTAGTAGACAGATTTGCCATGTTGACCAAGCTTG 1432  
|||  
DB 132722 CACCCGGCTAATTTTGTATTTTGTAGTAGACAGGAGTTTCACTATGTTGGCCAGGCTGG 132663  
|||  
QY 1433 CCTGAAGTCTGAGTCTGAGTATCCACCCAGCTCAGCTCCCAAGCACTGGGATTAC 1492  
|||  
DB 132662 TTGGAATCTCTGACTCCTCAGTGAITCCACCCACTTGGCTCCCAAGTGTGGATTAC 132603  
|||  
QY 1493 AGGATGAGCCACCGTGGCCAGCTGTTTCTCA 1526  
|||  
DB 132602 AGGATGAGCCACCGGCGCCGCTGTTTCTCA 132569  
|||

Search completed: January 22, 2006, 13:37:59  
Job time : 6825.68 secs



(b)(7)(D) - Exemption from disclosure



```
c 93 200.4 10.0 44075 12 ADM74369 Human car
c 94 200.4 10.0 47319 4 AAK72230 Human
c 95 200.4 10.0 47319 4 AAK64813 Human Imm
c 96 200.4 10.0 110000 13 ABD33780_3 Continuation (4 of
97 200.4 10.0 112132 6 ABK90888 Human ATP
98 200.4 10.0 112132 11 ADM56267 Human ATP
99 200.4 10.0 112132 14 ADV85874 Human pro
100 200.4 10.0 171158 12 ADQ97894 Adg97894 Human can
```

## ALIGNMENTS

## RESULT 1

AAV83939 standard; DNA; 80595 BP.

AAV83939;

03-MAR-1999 (first entry)

HC-Contig derived from normal human chromosome 10q25.2 region.

Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;  
neocentromere; replication; extra-chromosomal element; segregation;  
cell division; artificial chromosome; gene therapy; mardel(10);  
human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.

Homo sapiens.

WO9851790-A1.

19-NOV-1998.

13-MAY-1998; 98WO-AU000352.

13-MAY-1997; 97AU-00006784.

26-AUG-1997; 97AU-00008791.

(AMRA-) AMRAD OPERATIONS PTY LTD.

Choo K, Du Sart D, Cancellia MR;

WPI; 1999-009773/01.

New isolated nucleic acid comprising neocentromere sequences from  
eukaryotic chromosome - used to produce replicable, segregating  
artificial chromosomes that can carry large amounts of DNA for gene  
therapy.

Claim 8; Fig 6; 540pp; English.

The present sequence represents the HC-contig derived from normal human  
chromosome 10, 10q25.2 region. This region can be naturally mutated to  
produce an unusual chromosomal marker referred to as mardel(10). The  
mardel(10) marker is mitotically stable and contains a functional  
neocentromere at a location regarded as non-centromeric. This  
neocentromere maps to q25.2 on chromosome 10. The specification describes  
nucleic acid sequences derived from a eukaryotic chromosome, including a  
neocentromere or its functional derivative or hybrid, that are able, in a  
compatible cell, of replicating, acting as extra-chromosomal element and  
segregating during cell division. The sequences can be used to construct  
artificial chromosomes for use in gene therapy comprising a replicable,  
segregating nucleic acid that confers a specific phenotype on cells.  
Human artificial chromosomes can propagate in human cells and carry large  
amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,  
they are not mutagenic. The artificial chromosomes are also useful for  
generation of transgenic plants and animals, in production of proteins  
and to make diagnostic reagents, e.g. for expression of cytokines,  
receptors and growth factors, or to increase the copy number of a gene in  
a cell. The constructs may also be used for functional and structural  
analysis of chromosomes

SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

Query Match 100.0%; Score 2001; DB 2; Length 80595;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AATGTTGGATTTCAGAGTAAAGTAAAGAGCTGGAAACAACTACTGCGCCGAGTCTGAAG 60
DB 19000 AATGTTGGATTTCAGAGTAAAGTAAAGAGCTGGAAACAACTACTGCGCCGAGTCTGAAG 19059
QY 61 TTTGAGCGGAGACGTGGTGCAGCTGTGTTTTCACGTGACAGCTGAATATGAGCCGAC 120
DB 19060 TTTGAGCGGAGACGTGGTGCAGCTGTGTTTTCACGTGACAGCTGAATATGAGCCGAC 19119
QY 121 TTCAGTGAAGCTGTGTTTCCCTCCCTCAAGGTTTCCACAAATTCAGATTTCTCTCAGG 180
DB 19120 TTCAGTGAAGCTGTGTTTCCCTCCCTCAAGGTTTCCACAAATTCAGATTTCTCTCAGG 19179
QY 181 AAAGCCAAAATATGAATTTGAGGGTTTGAATTTGTTCTTTTATTTACAGGATTTG 240
DB 19180 AAAGCCAAAATATGAATTTGAGGGTTTGAATTTGTTCTTTTATTTACAGGATTTG 19239
QY 241 AATAATGTTCTCCACAGAGTTCTGCTTGTAAACAATGCACTTCTGACACTACTG 300
DB 19240 AATAATGTTCTCCACAGAGTTCTGCTTGTAAACAATGCACTTCTGACACTACTG 19299
QY 301 CATATGACGAGGTGTTTACTCAAGTAAACAGAAATGAGCTGCGCCAAATTCAAATCC 360
DB 19300 CATATGACGAGGTGTTTACTCAAGTAAACAGAAATGAGCTGCGCCAAATTCAAATCC 19359
QY 361 TGAATCTGATGAGAGAAATCAGAAATTAATTAATGAGGATTTCAACAGAGCTGCTAGAGATG 420
DB 19360 TGAATCTGATGAGAGAAATCAGAAATTAATTAATGAGGATTTCAACAGAGCTGCTAGAGATG 19419
QY 421 TGCCAGTGTGATGATACCTTGTCTCATCTCATAGCAGAGTGTGCTGTAGCAACTGCTCA 480
DB 19420 TGCCAGTGTGATGATACCTTGTCTCATCTCATAGCAGAGTGTGCTGTAGCAACTGCTCA 19479
QY 481 CTGCTTCATTTCTGCTGCTTGTGCTTAAATACGCTTTCACAGCTCAATGGCTTTCTT 540
DB 19480 CTGCTTCATTTCTGCTGCTTGTGCTTAAATACGCTTTCACAGCTCAATGGCTTTCTT 19539
QY 541 CCTCTGCGAGTCAAGCTTTCTTTGGGTCAAAACAGCAATGATTTCTTTAGATCACTGTGT 600
DB 19540 CCTCTGCGAGTCAAGCTTTCTTTGGGTCAAAACAGCAATGATTTCTTTAGATCACTGTGT 19599
QY 601 ACTCAAGGAGCTACAGACANTGGGCAATCCACTTCACTCTGGAAGAAACAAATTTTA 660
DB 19600 ACTCAAGGAGCTACAGACANTGGGCAATCCACTTCACTCTGGAAGAAACAAATTTTA 19659
QY 661 TGAAGCCAGAGTTCATAGTGTCTTGTAGGTTGTTGTCAAGCCAAAGCCCAAGCTT 720
DB 19660 TGAAGCCAGAGTTCATAGTGTCTTGTAGGTTGTTGTCAAGCCAAAGCCCAAGCTT 19719
QY 721 TGTGCTTCAAAATGAAATTTAGAGAGCTTCAGAAACAATCCAAATTTTCAATGAGCTCA 780
DB 19720 TGTGCTTCAAAATGAAATTTAGAGAGCTTCAGAAACAATCCAAATTTTCAATGAGCTCA 19779
QY 781 CCCAATGATTAAGAAACAATTTGCCATATCTCATATGCCACCTTTTCAGGTGGAGATG 840
DB 19780 CCCAATGATTAAGAAACAATTTGCCATATCTCATATGCCACCTTTTCAGGTGGAGATG 19839
QY 841 TGAATGCTGAATGAGGTCAACAGATTTGCCCAACCAACTTTTGCAGAAAGAGTGAAGCT 900
DB 19840 TGAATGCTGAATGAGGTCAACAGATTTGCCCAACCAACTTTTGCAGAAAGAGTGAAGCT 19899
QY 901 CTGACTGGGAGCCCTAAATATGCAAAAGTTGATGAGCTTTCATGCAAGATATGAACCC 960
DB 19900 CTGACTGGGAGCCCTAAATATGCAAAAGTTGATGAGCTTTCATGCAAGATATGAACCC 19959
QY 961 GGTATGATATATGCTAAAGAGTTGGCTTTATGTTTCTTATCTTCAAAACCTGGTATG 1020
DB 19960 GGTATGATATATGCTAAAGAGTTGGCTTTATGTTTCTTATCTTCAAAACCTGGTATG 20019
```

```
OY 1021 AATAGATATGCTGTGTTCCCTTTAAAAAATGTCACAAATTGATTTATGATGCTGTGAT 1080
   |||
   |||
   |||
Db 20020 AATAGATATGCTGTGTTCCCTTTAAAAAATGTCACAAATTGATTTATGATGCTGTGAT 20079
   |||
   |||
   |||
OY 1081 AGTAATCAACAGATATGCTGTGATGCTTCAAGAAATGCTTCAAGAAATGCTTCAAGATTTT 1140
   |||
   |||
   |||
Db 20080 AGTAATCAACAGATATGCTGTGATGCTTCAAGAAATGCTTCAAGAAATGCTTCAAGATTTT 20139
   |||
   |||
   |||
OY 1141 ACCCATATGTCACAAAGAGAGGCCATTTTCAAGTTGAAATTTGTTACAGAAATTTTCA 1200
   |||
   |||
   |||
Db 20140 ACCCATATGTCACAAAGAGAGGCCATTTTCAAGTTGAAATTTGTTACAGAAATTTTCA 20199
   |||
   |||
   |||
OY 1201 TATATGTTTCTCAGATCTTTTATTTTATTTTATTTTATTTTGAACAGATCTCACTTTGTCAC 1260
   |||
   |||
   |||
Db 20200 TATATGTTTCTCAGATCTTTTATTTTATTTTATTTTATTTTGAACAGATCTCACTTTGTCAC 20259
   |||
   |||
   |||
OY 1261 CCAGGCTGAGTATACAGTGTGCTGTGCTGTGCTCACTGTGACCTCTGCTCCAGGTTCA 1320
   |||
   |||
   |||
Db 20260 CCAGGCTGAGTATACAGTGTGCTGTGCTGTGCTCACTGTGACCTCTGCTCCAGGTTCA 20319
   |||
   |||
   |||
OY 1321 ACCGATTCCTCTGTACGCTTCCCGAATAGCTGGGATTAAGGGGCGATGACCAACCATGCGC 1380
   |||
   |||
   |||
Db 20320 ACCGATTCCTCTGTACGCTTCCCGAATAGCTGGGATTAAGGGGCGATGACCAACCATGCGC 20379
   |||
   |||
   |||
OY 1381 TAATTTTGTATTTTATAGAGACAGATTTGCGCATGTTGACAGGCTTGCTTGAAAC 1440
   |||
   |||
   |||
Db 20380 TAATTTTGTATTTTATAGAGACAGATTTGCGCATGTTGACAGGCTTGCTTGAAAC 20439
   |||
   |||
   |||
OY 1441 TCCTGATCTCAGATATCCACCCAGCTCCCAAGACATGGGATTTAAGAGCATTA 1500
   |||
   |||
   |||
Db 20440 TCCTGATCTCAGATATCCACCCAGCTCCCAAGACATGGGATTTAAGAGCATTA 20499
   |||
   |||
   |||
OY 1501 GGCACCGTCCAGAGCTGTTTCTCAGATCTGTGATTTGTTCTGAAGCCTTCATTGTA 1560
   |||
   |||
   |||
Db 20500 GGCACCGTCCAGAGCTGTTTCTCAGATCTGTGATTTGTTCTGAAGCCTTCATTGTA 20559
   |||
   |||
   |||
OY 1561 TCTTCTTATTTATTTTGAAGATATACCTAAGTAAAGTTTAAATCAATCAATTTCTT 1620
   |||
   |||
   |||
Db 20560 TCTTCTTATTTATTTTGAAGATATACCTAAGTAAAGTTTAAATCAATCAATTTCTT 20619
   |||
   |||
   |||
OY 1621 TGGAAATTCCTGTGCTTCTTATTTCTTATTTCTTCAAAAATATGTTCAATAGCTGATGT 1680
   |||
   |||
   |||
Db 20620 TGGAAATTCCTGTGCTTCTTATTTCTTATTTCTTCAAAAATATGTTCAATAGCTGATGT 20679
   |||
   |||
   |||
OY 1681 ATGTTTCTTTCAAAATTAATTCATTTCTATCTCAAGATTTATCTCATGCTTAATTTGTAT 1740
   |||
   |||
   |||
Db 20680 ATGTTTCTTTCAAAATTAATTCATTTCTATCTCAAGATTTATCTCATGCTTAATTTGTAT 20739
   |||
   |||
   |||
OY 1741 TGAATAGTTCATCTTCTGTCATCCAGTTTCTGTCTTATTTTCACTCAATCTAAGCTTAA 1800
   |||
   |||
   |||
Db 20740 TGAATAGTTCATCTTCTGTCATCCAGTTTCTGTCTTATTTTCACTCAATCTAAGCTTAA 20799
   |||
   |||
   |||
OY 1801 TGGCTATTAGATTAAGAGCTTGTAAACAGATCTTCTTCCAAATATGCTTTATCTTTTGAAC 1860
   |||
   |||
   |||
Db 20800 TGGCTATTAGATTAAGAGCTTGTAAACAGATCTTCTTCCAAATATGCTTTATCTTTTGAAC 20859
   |||
   |||
   |||
OY 1861 TGCATGCCAGTGAACAACTGTTAAGTTTGAATTTCTTCAATCAATTCACAGAAATATGC 1920
   |||
   |||
   |||
Db 20860 TGCATGCCAGTGAACAACTGTTAAGTTTGAATTTCTTCAATCAATTCACAGAAATATGC 20919
   |||
   |||
   |||
OY 1921 TGCATCTCTTCTTCTTGAAGAGCAATGCCCAAGACAGCATTTGTATGATATGTAACGTA 1980
   |||
   |||
   |||
Db 20920 TGCATCTCTTCTTCTTGAAGAGCAATGCCCAAGACAGCATTTGTATGATATGTAACGTA 20979
   |||
   |||
   |||
OY 1981 ACAGGAGCATGGGTGATAGC 2001
   |||
   |||
   |||
Db 20980 ACAGGAGCATGGGTGATAGC 21000
   |||
   |||
   |||
```

```
RESULT 2
AAV83940
ID AAV83940 standard; DNA; 80240 BP.
```

```
XX AAV83940;
AC
XX
AC 03-MAR-1999 (first entry)
DT
XX
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
XX neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
XX human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
XX 26-MAY-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancellia MR;
XX
DR WPI, 1999-009773/01.
XX
XX New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
XX Claim 9; Fig 16A; 540bp; English.
XX
XX The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC mitotically stable and contains a functional neocentromere at a location
CC regarded as non-centromeric. This neocentromere maps to q25.2 on
CC chromosome 10. The specification describes nucleic acid sequences derived
CC from a eukaryotic chromosome, including a neocentromere or its functional
CC derivative or hybrid, that are able, in a compatible cell, of
CC replicating, acting as extra-chromosomal element and segregating during
CC cell division. The sequences can be used to construct artificial
CC chromosomes for use in gene therapy comprising a replicable, segregating
CC nucleic acid that confers a specific phenotype on cells. Human artificial
CC chromosomes can propagate in human cells and carry large amounts of DNA
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC mutagenic. The artificial chromosomes are also useful for generation of
CC transgenic plants and animals, in production of proteins and to make
CC diagnostic reagents, e.g. for expression of cytokines, receptors and
CC growth factors, or to increase the copy number of a gene in a cell. The
CC constructs may also be used for functional and structural analysis of
CC chromosomes
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;
XX
Query Match 98.7%; Score 1974.2; DB 2; Length 80240;
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 3; Indels 2; Gaps 2;
Matches 1998; Conservative 0;
OY 1 AATGTTGGATTTCAAGATTAACAAGAGCTGGAAACAATCACTAGCGGAGTCTGAAG 60
   |||
   |||
   |||
Db 18736 AATGTTGGATTTCAAGATTAACAAGAGCTGGAAACAATCACTAGCGGAGTCTGAAG 18795
   |||
   |||
   |||
OY 61 TTTGACGGAGACATGCTGCTGCTGTTTCTTCACTGACAGCTGGAATAAGGCCACG 120
   |||
   |||
   |||
Db 18796 TTTGACGGAGACATGCTGCTGCTGTTTCTTCACTGACAGCTGGAATAAGGCCACG 18855
   |||
   |||
   |||
OY 121 TTCAGTGAAGCTTGTCTTCCCTCTCAAGATTTACCAACAATTTCTGATTTCTCAAG 180
   |||
   |||
   |||
```

Dh 18856 TTCAGTGAAGCTGTGTTCTTCCTCCCTCAAGTTACCCACAATTCGAGTTCTCTCAGG 18915  
Qy 181 AAGGCCAAAAATGAATTTGAGAGGTTTAGAGATTGTGTTCTTTATCTATTAACAGAAATTG 240  
Db 18916 AAGGCCAAAAATGAATTTGAGAGGTTTAGAGATTGTGTTCTTTATCTATTAACAGAAATTG 18975  
Qy 241 ATAATATGTTCTCCACAGATGTTCTGTTTAAACAATCTACGTTCTGACACTACTG 300  
Db 18976 ATAATATGTTCTCCACAGATGTTCTGTTTAAACAATCTACGTTCTGACACTACTG 19035  
Qy 301 CATATGCGAGATGTTACTACCAAGTAAACACAGAAATTGGCTGCCAATTCCAATCCC 360  
Db 19036 CATATGCGAGATGTTACTACCAAGTAAACACAGAAATTGGCTGCCAATTCCAATCCC 19095  
Qy 361 TGAATCGATGAGAAATCAGAAATTAATAGGGGATTCAACAGAGCTGCGCTAGAGATG 420  
Db 19096 TGAATCGATGAGAAATCAGAAATTAATAGGGGATTCAACAGAGCTGCGCTAGAGATG 19155  
Qy 421 TGCCAGTGGTCAGATTAATTGCTCATCATACGAGTGTGCTGCTCTAGCAACTGCTCA 480  
Db 19156 TGCCAGTGGTCAGATTAATTGCTCATCATACGAGTGTGCTGCTCTAGCAACTGCTCA 19215  
Qy 481 CTGCTTCATTTCTGCTGTTGTTCTTTAAATCTGCTTTTCTAGCTCAATTGGCTTTCTT 540  
Db 19216 CTGCTTCATTTCTGCTGTTGTTCTTTAAATCTGCTTTTCTAGCTCAATTGGCTTTCTT 19275  
Qy 541 CCTCTGGAGATCAGTTTCTTTGGGTCAAACGCAAAATGATTTCTTAGAATCACCTGGAT 600  
Db 19276 CCTCTGGAGATCAGTTTCTTTGGGTCAAACGCAAAATGATTTCTTAGAATCACCTGGAT 19335  
Qy 601 ACTCAAGAGAGCTACAGACATTTGGGATCAGCTTCCACTCTCTTGGAAAAACAATTTTA 660  
Db 19336 ACTCAAGAGAGCTACAGACATTTGGGATCAGCTTCCACTCTCTTGGAAAAACAATTTTA 19395  
Qy 661 TGGAAAGCCAGGTTGCTCATAGTGCCTCTTGAGGTTGTTGCTCAGCGCAAGGCCAGACTT 720  
Db 19396 TGGAAAGCCAGGTTGCTCATAGTGCCTCTTGAGGTTGTTGCTCAGCGCAAGGCCAGACTT 19455  
Qy 721 TGTGCTTCAACATGAATTTAGAGAGCTCAGAACAGATCCACATTTTCAATGGGCTCA 780  
Db 19456 TGTGCTTCAACATGAATTTAGAGAGCTCAGAACAGATCCACATTTTCAATGGGCTCA 19515  
Qy 781 CCCAATGAGTAAAGAACAAATTGCCATATCTCAATGACCACCTTTT -TCAGGTGGAGT 839  
Db 19516 CCCAATGAGTAAAGAACAAATTGCCATATCTCAATGACCACCTTTT -TCAGGTGGAGT 19575  
Qy 840 GTAGATGCTGGAATGGGTGACAGCATTCGCCAACAACCTTTGCAAAAAAGGCTGGAAAC 899  
Db 19576 GTAGATGCTGGAATGGGTGACAGCATTCGCCAACAACCTTTGCAAAAAAGGCTGGAAAC 19635  
Qy 900 TCTGACTGGGGACCCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAATATGAAGCC 959  
Db 19636 TCTGACTGGGGACCCCTAAATATGCAAAAGTTGATAGGCTCTTCATGCAATATGAAGCC 19695  
Qy 960 CGTGTATGATATAGCTAAAGGTTGGCTTTATGTTTCTATTCCTTCACAAACCTGATA 1019  
Db 19696 CGTGTATGATATAGCTAAAGGTTGGCTTTATGTTTCTATTCCTTCACAAACCTGATA 19755  
Qy 1020 GAATGATATGCTGTTTCCCTTTAAAAAATGTCACAACATGCAATTATGATGCTGTGTA 1079  
Db 19756 GAATGATATGCTGTTTCCCTTTAAAAAATGTCACAACATGCAATTATGATGCTGTGTA 19815  
Qy 1080 TAGTAACTCAGATATGCTGCATGAAAAATGCTTAGAAGCCCAATATATAGAGATTTT 1139  
Db 19816 TAGTAACTCAGATATGCTGCATGAAAAATGCTTAGAAGCCCAATATATAGAGATTTT 19875  
Qy 1140 TAGCCATGTGTGCAAAAGAGAGCCATTTCACTGTGGAATGTTCAAGAGATATTG 1199  
Db 19876 TAGCCATGTGTGCAAAAGAGAGCCATTTCACTGTGGAATGTTCAAGAGATATTG 19935  
Qy 1200 ATTATGTTTTCTCAGATCTTTTTTATTTTTTATTTTTTGAACAGAGTCTCACTTTGTCA 1259  
Db 19936 ATTATGTTTTCTCAGATCTTTTTTATTTTTTATTTTTTGAACAGAGTCTCACTTTGTCA 19995

Qy 1260 CCAGGCTGAGATACAGTGGCTGTGTCTGGCTCACTGCAACCTTGCTCCAGGTTTC 1319  
Db 19996 CCAGGCTGAGATACAGTGGCTGTGTCTGGCTCACTGCAACCTTGCTCCAGGTTTC 20055  
Qy 1320 AAGGATTTCTCTGCTCAGCTTCCCGAATAGCTGGGATTAACAGGCGGACACCAATGAC 1379  
Db 20056 AAGGATTTCTCTGCTCAGCTTCCCGAATAGCTGGGATTAACAGGCGGACACCAATGAC 20115  
Qy 1380 CTAATTTTTATTTTTATAGTAGACAGAGTTTGGCCATGTGTGACCAAGCTTGGCTTGA 1439  
Db 20116 CTAATTTTTATTTTTATAGTAGACAGAGTTTGGCCATGTGTGACCAAGCTTGGCTTGA 20175  
Qy 1440 CTCCTGACTTGAAGTATCCACCACTCAGCTTCCCAAGCACTGGGATTAACAGGATG 1499  
Db 20176 CTCCTGACTTGAAGTATCCACCACTCAGCTTCCCAAGCACTGGGATTAACAGGATG 20235  
Qy 1500 AGCCACGCTGCCAGCCTGTTTCTCAGATCTGTGTA -TTTGTTTCTGAAGCTTCATTTTC 1558  
Db 20236 AGCCACGCTGCCAGCCTGTTTCTCAGATCTGTGTA -TTTGTTTCTGAAGCTTCATTTTC 20295  
Qy 1559 TATCTTCTATTCATTTTGAAGTAGTACACTAAGGTTTAAACAATATATATC 1618  
Db 20296 TATCTTCTATTCATTTTGAAGTAGTACACTAAGGTTTAAACAATATATATC 20355  
Qy 1619 TTTGGAAAAATTCCTGCTGTTCTTTCTTATTTCTCAAAAAATATATGTTAGATAGCTGATG 1678  
Db 20356 TTTGGAAAAATTCCTGCTGTTCTTTCTTATTTCTCAAAAAATATATGTTAGATAGCTGATG 20415  
Qy 1679 TTATGTTTCTTTCAAATATATTCATTTCTCTATCTCAGAAATTTATCTCATGCTTATGTT 1738  
Db 20416 TTATGTTTCTTTCAAATATATTCATTTCTCTATCTCAGAAATTTATCTCATGCTTATGTT 20475  
Qy 1739 ATTGAATAGTCTTCACTTCTGTGCATCCAGTTTGTGCTCTTATTTCACTAAGTCTTA 1798  
Db 20476 ATTGAATAGTCTTCACTTCTGTGCATCCAGTTTGTGCTCTTATTTCACTAAGTCTTA 20535  
Qy 1799 AGTGGCTATTTGAATTAAGAGGCTTGTAAACAGATTTCTTCCAAATATGCTTATCTTTTG 1858  
Db 20536 AGTGGCTATTTGAATTAAGAGGCTTGTAAACAGATTTCTTCCAAATATGCTTATCTTTTG 20595  
Qy 1859 ACTGCATGCCAGTGCACAACTGTTAACGTTTGATTTCTTCAATCATTCACAGAAACT 1918  
Db 20596 ACTGCATGCCAGTGCACAACTGTTAACGTTTGATTTCTTCAATCATTCACAGAAACT 20655  
Qy 1919 GCTGACTGCTCTCTCTGAAAGCAATGCCAAGACAGCATTTGTAAGTAGATGTACG 1978  
Db 20656 GCTGACTGCTCTCTCTGAAAGCAATGCCAAGACAGCATTTGTAAGTAGATGTACG 20715  
Qy 1979 CAACAGGACATGGGTGCATAGC 2001  
Db 20716 CAACAGGACATGGGTGCATAGC 20738

RESULT 3  
ACN43882  
ID ACN43882 standard; DNA; 60815 BP.  
XX  
AC ACN43882;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human genomic sequence hCG177139.  
XX  
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2003073826-A2.  
XX  
XX 12-SEP-2003.  
PD 28-FEB-2003; 2003WO-US006235.  
PF

XX 01-MAR-2002; 2002US-00087192.  
 PR (SAGR-) SAGRES DISCOVERY.  
 XX  
 PA Morris DW;  
 XX  
 PI MPI; 2003-328604/31.  
 XX  
 DR Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 XX comprises a nucleotide sequence.  
 PT  
 PS Claim 1; SEQ ID NO 52; 0bp; English.  
 XX  
 CC The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent to basic patent  
 CC US2002182586A1, for which no sequence data was published  
 CC  
 XX  
 SQ Sequence 60815 BP; 14386 A; 15367 C; 17015 G; 13945 T; 0 U; 102 Other;  
 Query Match 10.8%; Score 215.8; DB 11; Length 60815;  
 Best Local Similarity 71.4%; Pred. No. 7e-44; Mismatches 122; Indels 3; Gaps 2;  
 Matches 312; Conservative 0; Mismatches 122; Indels 3; Gaps 2;  
 XX  
 YY 1102 CATGAATGCTTCAAGACCAATATTAAGAGATTTTATGACATGTGACAAAGAA 1161  
 DB 36062 CCTGATGACCTTGGACTCAGATCCAGGTCATTTTGGCCAAAGTGTACTACCTT 36121  
 YY 1162 GGCCTATTCAGTGTGAAATTTGTCAGAGATTTGATATGTTTCTCAGATCTTT 1221  
 DB 36122 TCCAGAGCAGTCTCCCAACCTCAATTTAGTTCTTCTCTCTCTCTCTCTTTT 36181  
 YY 1222 TATTTTATTTTGTGAAAGAGTCTACTTTGTACCCAGGCTGAGTACAGTGGT 1281  
 DB 36182 TTTTATTTTATTTTGAAGGAGAGTCTGTGTGACCCAGGCTGAGTACAGTGG-T 36240  
 YY 1282 GTGCTCTGGCTCAGTCAACCTGCTCCAGGTTCAAGGATTTCTCTG--TCAGCT 1339  
 DB 36241 GCAAGTCTGGCTCAGTCAACCTGCTCCAGGTTCAAGGATTTCTCTGAGCTC 36300  
 YY 1340 TCCCAATATGCTGGATTTACAGGCGCAGTCAACCACTTATTTTGTATTTTACT 1399  
 DB 36301 TCCCAATATGCTGGATTTACAGGCGCAGTCAACCACTTATTTTGTATTTTACT 36360  
 YY 1400 AGAGACAGAGTTTCCGCAATGTTGACCAAGCTTGGCTTGAATCTTCTGACTTCAAGTATTC 1459  
 DB 36361 AGAGACAGAGTTTCCGCAATGTTGACCAAGCTTGGCTTGAATCTTCTGACTTCAAGTATTC 36420  
 YY 1460 ACCCACTCAGGCTCCCAAGAGATTTACAGGATTTACAGGATTTACAGGATTTACAGGATTT 1519  
 DB 36421 GCCCGCTTACCTCCCAAGAGATTTACAGGATTTACAGGATTTACAGGATTTACAGGATTT 36480  
 YY 1520 TTTTCAATCTCTGAT 1536  
 DB 36481 CAATTGAGTTTCCGCT 36497

XX 20-MAY-2004 (first entry)  
 DT Human gene associated with low HDL-C PRPC.  
 XX  
 DE Human; ds; SNP; single nucleotide polymorphism;  
 XX high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;  
 KW coronary artery disease; gene.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX FH replace(63956,C)  
 XX FT /tag=a  
 XX FT /standard\_name="single nucleotide polymorphism"  
 XX  
 XX US2004043389-A1.  
 XX  
 XX 04-MAR-2004.  
 XX  
 XX 04-SEP-2002; 2002US-00235192.  
 XX  
 XX 04-SEP-2002; 2002US-00235192.  
 XX  
 XX (VITL-) VITIVITY INC.  
 XX  
 XX McCarthy J;  
 XX  
 XX MPI; 2004-214170/20.  
 XX  
 XX Determining whether a subject has, or is at risk of developing, an  
 PT abnormally low high density lipoprotein-C (HDL-C) level comprises  
 PT detecting an allelic variant of a polymorphic region from any of a set of  
 PT 27 genes.  
 XX  
 XX Disclosure; SEQ ID NO 47; 37pp; English.  
 XX  
 XX The invention relates to determining whether a subject has, or is at risk  
 CC of developing, an abnormally low high density lipoprotein-C (HDL-C) level  
 CC comprising determining whether the subject has an allelic variant of a  
 CC polymorphic region from any of 27 genes (alleles listed in Table 5 of the  
 CC specification). Also included are determining whether a male subject has,  
 CC or is at risk of developing, an abnormally low HDL-C level, comprising  
 CC determining whether the male subject has an allelic variant of a  
 CC polymorphic region listed in Table 5 which is associated with abnormally  
 CC low HDL-C levels in males, and determining whether a female subject has,  
 CC or is at risk of developing, an abnormally low HDL-C level, comprising  
 CC determining whether the female subject has an allelic variant of a  
 CC polymorphic region listed in Table 5 which is associated with abnormally  
 CC low HDL-C levels in females. The allelic variant in determining whether a  
 CC subject has, or is at risk of developing, an abnormally low HDL-C level  
 CC is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBA1  
 CC 1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The  
 CC allelic variant in determining whether a male subject has, or is at risk  
 CC of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,  
 CC or PRAG 1 CG, or their complements. The allelic variants are also COL5A2  
 CC 1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their  
 CC complements. The methods are useful for diagnosing (a predisposition to)  
 CC abnormally low levels of low high density lipoprotein-C (HDL-C) in a  
 CC subject. The methods are useful in diagnosing (a predisposition to) or  
 CC prognosticating diseases and disorders associated with abnormal lipid  
 CC levels such as vascular and metabolic diseases, e.g., coronary artery  
 CC disease. The present sequence is a human gene containing a SNP (single  
 CC nucleotide polymorphism) associated with low high density lipoprotein-C  
 CC (HDL-C) levels.  
 XX  
 XX  
 XX Sequence 182328 BP; 56638 A; 36576 C; 34943 G; 54171 T; 0 U; 0 Other;  
 SQ  
 Query Match 10.5%; Score 209.4; DB 12; Length 182328;  
 Best Local Similarity 78.2%; Pred. No. 4.8e-42;  
 Matches 290; Conservative 0; Mismatches 76; Indels 5; Gaps 3;  
 YY 1176 TGAATTTTCAAGAGATTTGATATGTTTCTCAGATCTTTTATTTTATTTT 1235





PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249246P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250391P.
PR	01-DEC-2000;	2000US-0251160P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251899P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
PI	WPI; 2001-541565/60.	
DR		
XX		
PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system cancers	
XX	and metastases.	
PS	Disclosure; SEQ ID NO 12963; 1701bp + Sequence Listing; English.	
XX		
CC	The invention relates to novel genes (ABA11004-ABA21534) and proteins	
CC	(ABA1678-ABA18001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful in	
CC	the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and	
CC	ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow,	
CC	breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune	
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic	
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,	
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)	
CC	cardiovascular disorders such as myocardial ischaemias; (d) wound healing	
CC	; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)	
CC	infectious diseases such as viral, bacterial, fungal and parasitic	
CC	infections. Note: The sequence data for this patent did not form part of	
CC	the printed specification, but was obtained in electronic format directly	

CC	from WIPo at ftp.wipo.int/pub/published_pct_sequences
XX	
SO	Sequence 1425 BP; 368 A; 344 C; 344 G; 369 T; 0 U; 0 Other;
	Query Match 10.4%; Score 208.6; DB 5; Length 1425;
	Best Local Similarity 82.6%; Pred. No. 1e-42;
	Matches 275; Conservative 0; Mismatches 54; Indels 4; Gaps 3;
OY	1220 TTATTTTTTTTGTGAAACAGAGCTCACTTTGTCAACCAGCGTAGTACAGTG 1279
Dd	1302 TTTTTTTTTTTTTTTTTTTGAAGTAAGTCACCTCTGTACCCAGACTGGAGTACAAG 1243
OY	1280 CTGTGCTGTGGCTCACTGCAACTCTGCCCTCCAGGTTCAAGCATTCCTTG -TCAG 1337
Dd	1242 AACAA-TCTGGGCTCATTGCAACTCCGCTCCAGGTTCAAGGATTCCTGTGCCTCAG 1184
OY	1338 CTTCCCGAATAAGCTGGGATTAACAGGCGCAG -CACCAACATGCTAATTTTGTATTTT 1396
Dd	1183 CCTTCCTAGAAGCTGGGATTAACAGGCAATGACCATGCCCGCTAATTTTGTATTTT 1124
OY	1397 AGTAGAGACAGAGTTTGCCATGTTGACCAAGGCTTGCTTGAATCCTGACTTCAGGTGA 1456
Dd	1123 AGTAGAGACAGAGTTTACCATATTGGCCAGGCTGTCTGAACTCTGACCTCAAGTGA 1064
OY	1457 TCACCCACCTCAGCCTCCCAAAGCACTGGGATTAACAGGATGAGCCACCGTCCAGCC 1516
Dd	1063 TCACCCAACTCAGCCTCCCAAAGCTGTGGGATTAATAGGATGAGCCACCGTCCAGCC 1004
OY	1517 TGTTTTCTCAGATCTGTATTTTGTTTGTGAAGC 1549
Dd	1003 GAATTTTTCAGATTTTGAAGGATTAATGCTGC 971
RESULT 6	
AAK70203/c	
ID	AAK70203 standard; DNA; 409 BP.
XX	
AC	AAK70203;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25015.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM	cytostatic; gene therapy; vaccine; metastasis; de.
XX	
OS	Homo sapiens.
XX	
CN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US001354.
XX	
PR	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205515P.
PR	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214886P.
PR	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.
PR	07-JUL-2000; 2000US-0216880P.
PR	11-JUL-2000; 2000US-0217487P.
PR	11-JUL-2000; 2000US-0217486P.
PR	14-JUL-2000; 2000US-0218280P.
PR	26-JUL-2000; 2000US-0220963P.
PR	26-JUL-2000; 2000US-0220964P.
PR	14-AUG-2000; 2000US-0224518P.
PR	

PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227183P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231245P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239355P.  
PR 13-OCT-2000; 2000US-0239357P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251898P.  
PR 08-DEC-2000; 2000US-0251909P.  
PR 11-DEC-2000; 2000US-0254977P.  
PR 05-JAN-2001; 2001US-0259678P.  
  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-483426/52.  
  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and metastas.  
Disclosure; SEQ ID NO 25015; 3071pp + Sequence Listing; English.  
  
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
amino acid sequences given in AAM62170 to AAM91921. (I) have cytostatic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For  
example, they may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of (I) by expressing inactive proteins or to  
supplement the patient's own production of (I). Additionally, (I)  
polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention

XX SQ Sequence 409 BP; 132 A; 78 C; 125 G; 74 T; 0 U; 0 Other:  
Query Match 10.4%; Score 208.2; DB 4; Length 409;  
Best Local Similarity 82.9%; Pred. No. 7.7e-43;  
Matches 286; Conservative 0; Mismatches 53; Indels 6; Gaps 4;  
QY 1201 TTATGTTTCTGAGCTTTTATTTTATTTTGAACAGACTTCACCTTTCAC 1260  
DB 407 TTTCTTTTCTTCTCTTTTCTTTTCTTTTGTGACAGAGTCTGCTGTTTCAT 348  
QY 1261 CCAGCTGAGTAGACAGTGGCTGTGCTGAGCTCAGTGCACCTGCTCCAGGTTCA 1320  
DB 347 CCAGCTGAGTAGACAGTGG- TGTGATCTGGGCTCATGCAACTC- GCTCCAGGTTCA 290  
QY 1321 AGGATTTCTCTG- TTAGCTTCCGAATAGCTGGATTAACAGGCGCATGCACCA- 1376  
DB 289 AGCAATTTCTCTGCTCAGCTTCCCAAGTAGCTGGGATTACAGGCCATGCCACACACC 230  
QY 1377 TGCCTAATTTTGTATTTTATTTAGTAGAGAGAGTTTCCGCAATTTGACAGGCTTGCCTT 1436  
DB 229 CGGCTAATTTTGTATTTTATTTTATTTAGTAGAGAGAGGTTTCCACCATGTTGGCCAGGCTGGTCTT 170  
QY 1437 GAATCTGACTTCAGAGTATCCACCCACCTCAGCTCCCAAGACATGGGATTTACAGGC 1496  
DB 169 GAATCTGACTTCAGAGTATCCGCCCCACCTGCGCTCCCAAGTGTGGGATTAACAGGC 110  
QY 1497 ATGAGCACCGTGGCCAGGCTGTTTCTGAGATCTGATTGTT 1541  
DB 109 AAGAGACACCATGCCAGGCTCTTTTCTTATATATCACTTCTT 65  
RESULT 7  
ID AAL06997  
ID AAL06997 standard; DNA; 31813 BP.  
XX AAL06997;  
AC AAL06997;  
XX 21-NOV-2001 (first entry)  
DT 21-NOV-2001 (first entry)  
XX Human reproductive system related antigen DNA SEQ ID NO: 9685.  
DE Human reproductive system related antigen; reproductive system disorder;  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO200155320-A2.  
PN WO200155320-A2.  
XX 02-AUG-2001.  
PD 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001339.  
PF 17-JAN-2001; 2001WO-US001339.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225457P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 14-AUG-2000; 2000US-0225799P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.

XX	08-NOV-2000;	2000US-0246523P.	PR
XX	08-NOV-2000;	2000US-0246524P.	PR
XX	08-NOV-2000;	2000US-0246525P.	PR
XX	08-NOV-2000;	2000US-0246526P.	PR
XX	08-NOV-2000;	2000US-0246527P.	PR
XX	08-NOV-2000;	2000US-0246528P.	PR
XX	08-NOV-2000;	2000US-0246529P.	PR
XX	08-NOV-2000;	2000US-0246601P.	PR
XX	08-NOV-2000;	2000US-0246610P.	PR
XX	08-NOV-2000;	2000US-0246611P.	PR
XX	08-NOV-2000;	2000US-0246613P.	PR
XX	17-NOV-2000;	2000US-0249207P.	PR
XX	17-NOV-2000;	2000US-0249208P.	PR
XX	17-NOV-2000;	2000US-0249209P.	PR
XX	17-NOV-2000;	2000US-0249210P.	PR
XX	17-NOV-2000;	2000US-0249211P.	PR
XX	17-NOV-2000;	2000US-0249212P.	PR
XX	17-NOV-2000;	2000US-0249213P.	PR
XX	17-NOV-2000;	2000US-0249214P.	PR
XX	17-NOV-2000;	2000US-0249215P.	PR
XX	17-NOV-2000;	2000US-0249216P.	PR
XX	17-NOV-2000;	2000US-0249217P.	PR
XX	17-NOV-2000;	2000US-0249218P.	PR
XX	17-NOV-2000;	2000US-0249224P.	PR
XX	17-NOV-2000;	2000US-0249245P.	PR
XX	17-NOV-2000;	2000US-0249264P.	PR
XX	17-NOV-2000;	2000US-0249265P.	PR
XX	17-NOV-2000;	2000US-0249297P.	PR
XX	17-NOV-2000;	2000US-0249299P.	PR
XX	17-NOV-2000;	2000US-0249300P.	PR
XX	01-DEC-2000;	2000US-0250160P.	PR
XX	01-DEC-2000;	2000US-0250391P.	PR
XX	05-DEC-2000;	2000US-0251030P.	PR
XX	05-DEC-2000;	2000US-0251988P.	PR
XX	05-DEC-2000;	2000US-0256719P.	PR
XX	06-DEC-2000;	2000US-0251479P.	PR
XX	08-DEC-2000;	2000US-0251856P.	PR
XX	08-DEC-2000;	2000US-0251868P.	PR
XX	08-DEC-2000;	2000US-0251869P.	PR
XX	08-DEC-2000;	2000US-0251989P.	PR
XX	08-DEC-2000;	2000US-0251990P.	PR
XX	11-DEC-2000;	2000US-0254097P.	PR
XX	05-JAN-2001;	2001US-0259678P.	PR
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
DR	WPI; 2001-465570/50.		
XX			
PT	Isolated nucleic acid molecule encoding a reproductive system antigen is		
XX	used in preventing, treating or ameliorating a medical condition.		
XX			
PS	Disclosure; SEQ ID NO 9685; 1297bp + Sequence Listing; English.		
XX			
CC	The present invention provides the protein and coding sequences of a		
CC	number of human reproductive system related antigens. These can be used		
CC	in the prevention and treatment of reproductive system disorders,		
CC	including cancer. The present sequence is a genomic sequence encoding a		
XX	protein of the invention		
XX			
XX	Sequence 31813 BP; 7330 A; 7632 C; 7753 G; 9098 T; 0 U; 0 Other;		

	Query March	Similarity	Score	DB	Length
	Best local	78.0%	Pred. No. 5.3e-42		
	Matches 276	Conservative	0	Mismatches 75	Indels 3; Gaps 2;
Qy	1216	TCCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTCCACCAGCGGTGAGTACA	1275		
Db	21355	TTTTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTCTTGCTCTGTCAACCAGCGTGGAGTCA	2141		
Qy	1276	GTGGCTGTGTTCTGGGCTCACTGCACCTCTGGCTCCAGGTTGACGAGATTCTTCCTG--	1333		

Db	21415	GTGG-1GTATATCTACAGTCAAGGCAACCTCTGCTCCAGGTTCAAGTATCTCCGCG	2147
QY	1334	TCAGCTTCCCGAATACTCTGGGATTTACAGCGGATGACACACATGCTTAATTTTGATTT	1393
Db	21474	TCAGCTCCCAATACTCTGGGATTTACAGGCAAGTGACACACACGCGGTAATTTTGATTT	21533
QY	1394	TTTATAGTAGAGACAGATTTTGGCATGTTACACAGGCTTGCTTGAATCTCTGACTTCAGG	1453
Db	21534	TTTATAGTAGAGACAGGTTTGGCATGTTACAGGCTTGCTTGAATCTCTGACTTCAGG	21593
QY	1454	TGATTCACCCACTGACCTCTCCCAAGCACTGGGATTTACAGCATGAGCCACCGGCCCA	1513
Db	21594	TGATTCACCCACTGCGGCTCTCCCAAGTGCTGGGATTTACAGGCATGAGCCACTGCTCCCG	21653
QY	1514	GCCGTGTTTCTCAGATCTCGTATTTGTTCTGTAAGCTTCATTTGACTTCTT	1567
Db	21654	GCCCGGAATTTCTTTATTTCTATACATTTTACTTCCGTTGTTTCTTTT	21707
RESULT 8			
ABA08137	ID	ABA08137 standard, DNA; 31813 BP.	
XX	AC	ABA08137;	
XX	DT	11-JAN-2002 (first entry)	
DE	XX	Human ovarian and breast cancer associated polymnucleotide SEQ ID NO 932.	
KW	XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW	XX	antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;	
KW	XX	nutraceutical; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KW	XX	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;	
KW	XX	neurological disease; infection; human; secreted protein; ds.	
OS	XX	Homo sapiens.	
XX	XX	MO200155325-A2.	
PN	XX	02-AUG-2001.	
PD	XX	17-JAN-2001; 2001WO-US001345.	
XX	XX		
PR	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-0190076P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-020515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		
PR	30-JUN-2000; 2000US-0215135P.		
PR	07-JUL-2000; 2000US-0216647P.		
PR	07-JUL-2000; 2000US-0216880P.		
PR	11-JUL-2000; 2000US-0217487P.		
PR	11-JUL-2000; 2000US-0217496P.		
PR	14-JUL-2000; 2000US-0218290P.		
PR	26-JUL-2000; 2000US-0220963P.		
PR	26-JUL-2000; 2000US-0220964P.		
PR	14-AUG-2000; 2000US-0224518P.		
PR	14-AUG-2000; 2000US-0224519P.		
PR	14-AUG-2000; 2000US-0225213P.		
PR	14-AUG-2000; 2000US-0225214P.		
PR	14-AUG-2000; 2000US-0225266P.		
PR	14-AUG-2000; 2000US-0225267P.		
PR	14-AUG-2000; 2000US-0225268P.		
PR	14-AUG-2000; 2000US-0225270P.		
PR	14-AUG-2000; 2000US-0225447P.		
PR	14-AUG-2000; 2000US-0225757P.		
PR	14-AUG-2000; 2000US-0225758P.		
PR	14-AUG-2000; 2000US-0225759P.		









KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW Cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX 09-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001354.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225578P.  
PR 14-AUG-2000; 2000US-022558P.  
PR 14-AUG-2000; 2000US-022559P.  
PR 18-AUG-2000; 2000US-022679P.  
PR 22-AUG-2000; 2000US-0226881P.  
PR 22-AUG-2000; 2000US-022688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-023144P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251866P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.

PR 05-JAN-2001; 2001US-0259678P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM,  
 XX WPI; 2001-483426/52.  
 XX  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
 XX  
 XX Disclosure; SEQ ID NO 30489; 3071bp + Sequence Listing; English.  
 XX  
 XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 SO Sequence 16424 BP; 4399 A; 4136 C; 3644 G; 4245 T; 0 U; 0 Other;

Query Match 10.3%; Score 206.8; DB 4; Length 16424;  
 Best Local Similarity 74.6%; Pred. No. 8.1e-42;  
 Matches 300; Conservative 0; Mismatches 97; Indels 5; Gaps 3;

QY 1166 ATTGAGTGTGAANTGTTGAGAGATTTGATTGATTTTTCAGATCTTTTAA 1225  
 DB 9024 AATTCTGTGTATCTTCTAGAGATTAATCTTAATAAGATGAGACATATTTCTTT 9083  
 QY 1226 TTTATTTTGTGAAACAGAGTCTCACTTTGTCAACCAGCTGAGTACAGTGTG 1285  
 DB 9084 TTTATTTTGTGAGCAGAGTCTTGTCTGTCAACCAGCTGAGTACAGTGTG 9142  
 QY 1286 TCTGGCTCAGTCAACTCTGCTCCAGGTTCAAGCATTTCTCTGTC--AGCTTC 1343  
 DB 9143 TCTCAGCTCAGCAACTCTGCTCCAGGTTCAAGCATTTCTCTGTTAGCTCTC 9202  
 QY 1344 GAATGCTGGGATTCAGGCG--CANGCACACCAACCTATTTTGTATTTTAA 1401  
 DB 9203 GATGCTGGGATTCAGGCGCTCACACCAACCTGATTTTGTATTTTAA 9262  
 QY 1402 AGACAGAGTTTCCGATGTTGACAGGCTTGCCTTGAATCTCTGATCTGAGTGA 1461  
 DB 9263 TTAACAGGTTTACCATGTTGGCCAGGCTGTGCTTGAATCTCTGATCTGAGTGA 9322  
 QY 1462 CCACCTCAGCTCCCAAGCACTGGGATTACAGGATGAGCCACCGTCCAGCTGT 1521  
 DB 9323 CTGCTTGGCTCCCAAGATTAATGAGGATTAAGGCGTGAAGCATCTGCTCCAGTCTTT 9382  
 QY 1522 TCTCAGATCTGATTTTGTCTGAAGCTTCAATTTCTATCT 1563  
 DB 9383 TTTTGT 9424

RESULT 13  
 AAK68448/C  
 ID AAK68448 standard; DNA; 16424 BP.  
 AC AAK68448;  
 XX  
 DT 06-NOV-2001 (first entry)

XX  
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23260.  
 XX  
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytosolic; gene therapy; vaccine; metastasis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157182-A2.  
 PD  
 XX 09-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US001354.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-020515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218299P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225477P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226861P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 23-AUG-2000; 2000US-0227182P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0228927P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0232981P.  
 PR 14-SEP-2000; 2000US-0233397P.  
 PR 14-SEP-2000; 2000US-0233398P.  
 PR 14-SEP-2000; 2000US-0233399P.  
 PR 14-SEP-2000; 2000US-0234000P.  
 PR 14-SEP-2000; 2000US-0234010P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.

PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.

PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI, 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclousre; SEQ ID NO 23260; 3071bp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostratic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 16424 BP; 4245 A; 3644 C; 4136 G; 4399 T; 0 U; 0 Other;  
Query Match 10.3%; Score 206.8; DB 4; Length 16424;  
Best Local Similarity 74.6%; Pred. No. 8.1e-42;  
Matches 300; Conservative 0; Mismatches 97; Indels 5; Gaps 3;  
QY 1166 ATTTCAGTGTGAATTTGTCAGAGATATTTGATATTTGTCAGATCTTTTAT 1225  
DB 7401 AATTCGTGTGTATCTTCTAGATATATCTATATTAAGATGAACATATTTCTTT 7342  
QY 1226 TTTATTTTTTTTGAACAAGAGTCTCATCTTGTACCCAGGCTGAGTACAGTGGCTGG 1285  
DB 7341 TTTATTTTTTTTGAACAAGAGTCTGCTGTGTACCCAGGCTGAGTACAGTGGC-ACAA 7283  
QY 1286 TCTGGCTACTGCAACTTGTGCTTCCAGAGTTCAAGGATTTCTCTGC-AGCTTCCC 1343  
DB 7282 TCTCAGCTCAGTCAACCTGTGCTTCCAGAGTTCAAGGATTTCTCTTGAAGCTCTGC 7223  
QY 1344 GAATGCTGCGATTACAGGCG--CATGACCAACCATGCTAATTTTGATTTTAGTAG 1401  
DB 7222 GAGTAGCTGAATTAAGAGCGTCACACACAGCGCTGATATTTTGTATTTTAAACAG 7163  
QY 1402 AGACAGAGTTTGGCCATGTTGACAGGCTGTGCTTGAATCTCTGACTTCAGGTGATTCAC 1461  
DB 7162 TTACAGGGTTTACCATGTGTGCGCAGGCTGTGATCTCTGACTGACCTGAGTATTCAC 7103  
QY 1462 CCACCTCAGCCTCCCAAGACCTGAGATTTACAGCATAGACCAAGCGTCCAGCTGTTT 1521  
DB 7102 CTGCTTGCGCTCCCAAGATTAAGGATTTACAGGCGTGAAGCAGCATGTGCCAGTCTTTT 7043  
QY 1522 TCTCAGATCTGATATTTGTTCTGAAGCTTATTTTATCT 1563  
DB 7042 TTTTTTTTTTTTTTTTTTTTTTGAAGTGTATATCTCTGT 7001  
RESULT 14  
ADP75184/c  
ID ADP75184 standard; DNA; 35192 BP.  
XX

AC ADP75184;  
 XX 12-AUG-2004 (first entry)  
 XX Human ADAM19 gene fragment #4.  
 XX  
 XX Human; chromosome 5; ds; gene; ADAM19; Endophilin 1; Endophilin 2; NRG2;  
 KW ADAMTS2; a disintegrin and metalloprotease; neuroligin 2; SNP;  
 KW single nucleotide polymorphism;  
 KW a disintegrin and metalloprotease with thrombospondin type1 motif 2;  
 KW asthma; atopy; obesity; inflammatory bowel disease; respiratory disorder.  
 OS Homo sapiens.  
 XX WO2003031594-A2.  
 XX  
 XX 17-APR-2003.  
 PD  
 XX 11-OCT-2002; 2002WO-US032700.  
 PF  
 XX 11-OCT-2001; 2001US-0328424P.  
 PR  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX Keith T, Little RD, Van Berdewegh P, Dupuis J, Del Maestro RG,  
 PI Allen K,  
 XX  
 DR WPI; 2003-381712/36.  
 PT New isolated nucleic acid or alternate splice variant, useful for  
 PT diagnosing and treating a disintegrin and metalloprotease (ADAM) or  
 PT interactor gene-associated disorder, e.g. asthma, atopy, obesity or  
 PT inflammatory bowel disease.  
 PS  
 PS Claim 2; SEQ ID NO 5; 338pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid or alternate splice  
 XX variant comprising a nucleotide sequence containing at least one of the  
 XX single nucleotide polymorphisms given in the specification, a nucleotide  
 XX sequence having at least 15 contiguous nucleotides of them, or  
 XX complements of them. The genes are ADAM19 (a disintegrin and  
 XX metalloprotease 19, also known as gene 845), NRG2 (neuregulin 2, also  
 XX known as gene 847), endophilin 1 (also known as gene 874), endophilin 2  
 XX (also known as gene 803) and ADAMTS2 (a disintegrin and metalloprotease  
 XX with thrombospondin type1 motif 2, also known as gene 962). Also included  
 XX are a vector comprising the isolated nucleic acid (or alternate splice  
 XX variant), a host cell containing the isolated nucleic acid (or alternate splice  
 XX encoded by the novel nucleic acid (or alternate splice variant), an  
 XX antibody or antibody fragment that binds to the polypeptide, an  
 XX pharmaceutical compositions (comprising the nucleic acid or alternate  
 XX splice variant, vector, polypeptide or antibody, and a carrier,  
 XX excipient or diluent), a kit for detecting a disintegrin and  
 XX metalloprotease (ADAM) gene nucleotide sequence (comprising the isolated  
 XX nucleic acid or alternate splice variant, antibody or antibody fragment,  
 XX and at least one component to detect the hybridisation of the variant or  
 XX the binding of the antibody to an ADAM gene amino acid sequence), a kit  
 XX for detecting an interactor gene amino acid sequence (comprising the  
 XX antibody or antibody fragment, and at least one component to detect the  
 XX binding of the antibody to the interactor gene amino acid sequence),  
 XX diagnosing an ADAM or interactor gene-associated disorder or a  
 XX respiratory disorder in a human subject, determining an ADAM or  
 XX interactor gene pharmacogenetic profile in a human subject, identifying  
 XX an orthologue of a human ADAM or interactor gene, treating an ADAM or  
 XX interactor gene-associated disorder (or a respiratory disorder) by  
 XX administering the pharmaceutical composition, a transgenic mouse (whose  
 XX genome comprises an introduced null mutation in an endogenous gene that  
 XX is orthologous to a human ADAM gene), making a homozygous transgenic  
 XX knockout mouse, forming a crystal of the isolated polypeptide, a cell  
 XX line comprising the isolated nucleic acid or alternate splice variant, a cell  
 XX biochip comprising the isolated nucleic acid or alternate splice variant,  
 XX an isolated nucleic acid probe or primer comprising at least 8 contiguous  
 XX nucleotides of the nucleic acid, an isolated antisense nucleic acid,  
 XX identifying an ADAM or interactor gene ligand and an isolated nucleic

CC acid variant of Gene 803, 845, 847, 874 or 962. The nucleic acid or  
 CC alternate splice variants, methods, kits and antibody/antibody fragment  
 CC are useful for diagnosing and treating an ADAM or interactor gene-  
 CC associated disorder, e.g. asthma, atopy, obesity or inflammatory bowel  
 CC disease. The present sequence is a gene (or gene fragment) for one of the  
 CC above mentioned genes. NOTE: This sequence may contain one of the SNPs  
 CC detailed in table 5 of the specification but their positions cannot be  
 CC determined using the information given in the patent.  
 XX  
 XX Sequence 35192 BP; 9087 A; 7990 C; 8613 G; 9502 T; 0 U; 0 Other;  
 SO  
 Query Match 10.3%; Score 206.8; DB 11; Length 35192;  
 Best Local Similarity 74.6%; Pred. No. 1.1e-41;  
 Matches 300; Conservative 0; Mismatches 97; Indels 5; Gaps 3;  
 QY 1166 ATTTGAGTGTGAATTGTTGAGAGAGATTTGATTTGTTCTGAGATCTTTTATTT 1225  
 DB 8751 AATTCGTGTGTATCTCTAGAGATTAATCTATTAAGAAAGATGATATTTCTTTT 8692  
 QY 1226 TTTATTTTGTGAAGAGAGCTTCATTTGTCAACCGGCTGAGTACAGTGGCTGTGG 1285  
 DB 8691 TTTATTTTGTGACACAGAGCTTGTCTGTACCCAGGCTGAAATGAGTGGC-ACAA 8633  
 QY 1286 TCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTCCTGTC--AGCTTCCC 1343  
 DB 8632 TCTGAGCTCACGCAACCTCTGCTCCAGGTTCAAGGATTCCTGTCAGCTTC 8573  
 QY 1344 GAATGCTGGGATTAACAGCGG--CATGCAACCAATGCTTAATTTTGTATTTAGTAG 1401  
 DB 8572 GAGTAGCTGAATTAACAGGCGTCCACCAACGCTGCTAATTTTGTATTTAGAGC 8513  
 QY 1402 AGACGAGTTTGGCATGTTGACCAAGGCTTGTGTTGAACTCTGATCTTCAAGTATCCAC 1461  
 DB 8512 TTACAGGGTTTCACATGTGTGCGCAGGCTGTGAACTCTGACCTCAGGTGATTCAC 8453  
 QY 1462 CCACCTGAGCTCCCAAGACACTGAGATTACAGCATGAGCCACGCTGAGCTGTTT 1521  
 DB 8452 CTGCTTGGCTCTCCCAAGTACTGAGATTACAGGCTGAGCCACGCTGCTTTT 8393  
 QY 1522 TCTCAGATCTGTATTTGTTTCTGAAGCTTCATTTTATCTT 1563  
 DB 8392 TTTTGT 8351  
 RESULT 15  
 ID ABD32861/C  
 AB032861 standard; DNA; 118466 BP.  
 AC ABD32861;  
 XX  
 XX 18-NOV-2004 (first entry)  
 DT  
 XX Human cancer-associated genomic DNA HD17-042.  
 DE  
 XX Human; ds; cancer-associated protein; gene; cytosolic; cancer;  
 KW leukaemia; lymphoma; CAP.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO2004074320-A2.  
 PN  
 XX  
 XX 02-SEP-2004.  
 PD  
 XX  
 XX 17-FEB-2004; 2004WO-US004730.  
 PF  
 XX  
 XX 14-FEB-2003; 2003US-00367094.  
 PR 14-MAR-2003; 2003US-00388938.  
 PR 15-APR-2003; 2003US-00417375.  
 PR 13-JUN-2003; 2003US-00461862.  
 PR 15-SEP-2003; 2003US-00663431.  
 PR 15-DEC-2003; 2003US-00737318.  
 XX  
 XX (SAGR-) SAGRES DISCOVERY INC.

XX PI Morris DW, Morris DW, Malandro MS;  
XX DR WPI; 2004-652914/63.  
XX PT New isolated cancer-associated polynucleotides and polypeptides useful  
XX PR for diagnosing, preventing or treating cancers, especially lymphoma and  
XX leukemia, or in screening for agents that modulate cancer.  
PS claim 16; seqid 536; 310pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising at least 10  
XX CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
XX CC in the specification, or its complement. The nucleic acids encode cancer-  
XX CC associated proteins. Also included are an expression vector comprising  
XX CC the isolated nucleic acid cited above, a host cell comprising the above  
XX CC recombinant nucleic acid or expression vector, a microarray for detecting  
XX CC a cancer-associated (CA) nucleic acid comprising at least one probe  
XX CC comprising at least 10 contiguous nucleotides of any of the above-  
XX CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
XX CC an open reading frame of a CA sequence selected from any of the 95  
XX CC polynucleotide sequences as mentioned in the specification, or its  
XX CC complement), an isolated antibody, (or its antigen binding fragment) that  
XX CC binds to the above polypeptide, a hybridoma that produces the above  
XX CC monoclonal antibody, a pharmaceutical composition comprising the above  
XX CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
XX CC cells (comprising the antibody cited above, methods for diagnosing cancer  
XX CC or for detecting the presence or absence of cancer cells in an  
XX CC individual, a method for inhibiting growth of cancer cells in an  
XX CC individual, a method for delivering a therapeutic agent to cancer cells  
XX CC in an individual, an electronic library comprising the above  
XX CC polynucleotide or polypeptide (or their fragments), methods of screening  
XX CC for anticancer activity or for a bioactive agent capable of modulating  
XX CC the activity of a CA protein (CAP), methods for detecting cancer  
XX CC associated with expression of a polypeptide in a test cell sample, a  
XX CC method for treating cancers and a method for inhibiting the expression of  
XX CC CA gene in a cell. The composition and methods are useful for detecting,  
XX CC diagnosing, preventing and treating cancers, especially lymphoma and  
XX CC leukemia. These may also be used in screening for agents that modulate  
XX CC cancer. The present sequence is a human CAP genomic sequence. Note: The  
XX CC sequence data for this patent did not form part of the printed from WIPO  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences  
SQ Sequence 118466 BP; 30920 A; 25969 C; 28235 G; 33263 T; 0 U; 79 Other;  
Query Match 10.3%; Score 206.8; DB 13; Length 118466;  
Best Local Similarity 74.6%; Pred. No. 1.8e-41;  
Matches 300; Conservative 0; Mismatches 97; Indels 5; Gaps 3;  
QY 1166 ATTTGAGTGTGAATGTTGTCAGAGATGATTGATTGTTTTCAGATCTTTTAT 1225  
DB 90189 AATTCCTGCTGATCTTCTTAGAGATATATCTATTAAGATGAGACATATTTT 90130  
QY 1226 TTTATTTTTTGAACAGAGTCTCACTTTGTACACCAGCTGAGATCAAGTGGCTG 1285  
DB 90129 TTTATTTTTTGAAGACAGAGCTTGTCTGTACACCAGCTGAGATGAGTGGC-ACAA 90071  
QY 1286 TCTGGGCTACATGCAACCTCTGCTCCAGGTTCAAGCATTTCTCTGTC--AGCTTCCC 1343  
DB 90070 TCTCAGCTCAGTCAACCTCTGCTCCAGTTTCAAGATTTCTCTGCTTAGCCTTTC 90011  
QY 1344 GAATGCTGTGGATTAACAGGCG--CATGACCAACCAAGCTAATTTTGTATTTTATGAG 1401  
DB 90010 GAGTACGTGAATTAACAGCGCTCCACCAACGCGCTGAATTTTGTATTTTATGAG 89951  
QY 1402 AGACAGAGTTTGGCCATGTTGACCAAGCTTGCTTGAATCTCTGATCAGTGATCAAC 1461  
DB 89950 TTACAGAGGTTTACCATGTTGGCCAGCTGTGTTTAATCTCTGATCAGTGATCAAC 89891  
QY 1462 CCACTCAGCTCCCAAGACACTGGAGTTAAGAGCATGAGCCACGCTGCCACCTGTTT 1521  
DB 89890 CTGCTTGGCTCCCAAGATGAGGATTAAGAGCGCTGAGCCACTGTGCCAGTCTTTT 89831

QY 1522 TCTCAGATCCTGTAATTTGTTTCTGAAGCTTCATTTCTATCT 1563  
DB 89830 TTTTATTTTTTTTTTTTTTTTGAAGATGATCATCCCTCTGT 89789  
RESULT 16  
ACNA3946  
ID ACNA3946 standard; DNA; 174448 BP.  
XX ACNA3946;  
XX AC 18-NOV-2004 (first entry)  
XX DT Human genomic sequence hCG21793.  
XX DE  
XX DE Human genomic sequence hCG21793.  
XX DE  
XX DE Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX KW  
XX OS Homo sapiens.  
XX PN WO2003073826-A2.  
XX PD 12-SEP-2003.  
XX PF 28-FEB-2003; 2003WO-US006235.  
XX PR 01-MAR-2002; 2002US-00087192.  
XX PA (SAGR-) SAGRES DISCOVERY.  
XX PI Morris DW;  
XX DR WPI; 2003-328604/31.  
XX PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX PT comprises a nucleotide sequence.  
PS Claim 1; SEQ ID NO 148; 0pp; English.  
XX CC The present invention relates to novel DNA and protein sequences which  
XX CC are associated with carcinomas. The sequences are useful for: (i) for  
XX CC screening drug candidates; (ii) for screening of bioactive agent capable  
XX CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
XX CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
XX CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX CC carcinoma including lymphoma. The present sequence is one such CA coding  
XX CC sequence. Note: This patent is an equivalent to basic patent  
XX CC US2002182586A1, for which no sequence data was published  
SQ Sequence 174448 BP; 43106 A; 40529 C; 40481 G; 47948 T; 0 U; 2384 Other;  
Query Match 10.3%; Score 206.8; DB 11; Length 174448;  
Best Local Similarity 79.4%; Pred. No. 2.2e-41;  
Matches 270; Conservative 0; Mismatches 67; Indels 3; Gaps 2;  
QY 1180 ATTTGTCAGAGAAATGATTGATTGTTTCTCAATCTTTTATTTTATTTTATTTTGA 1239  
DB 115868 AATGATGATATACATTAATGTTTAAATATATTAATTAATTTTCTTTTATTTTGA 115927  
QY 1240 AACAGAGTTCATCTTTGTACCCAGAGCTGAGATACAGGCTGTGCTCCGCTACATGC 1299  
DB 115928 CACGAGTCTCACTCCGTACCCAGAGCTGAGTGAAGTGGCT-TCATCTTGGCTCACTGC 115986  
QY 1300 AACCTGTGCTCCCAAGGTTCAAGCATTTCTCTG--TCAGTTTCCGAATAGCTGGGATT 1357  
DB 115987 AACCTTGGCTCTGTGGGTTCAAGCATTTTGTGCTTATGCTCTCGAATAGCTGGATT 116046  
QY 1358 ACAGGCGATGACCAACCAATGCTAATTTTGTATTTTATGATGAGACAGAGTTTCGCA 1417

Db 116047 ACAGGTGCTGCACATACACCGGCTATTTTGTATTTAGTAGAGAGCGGCTTTCACCA 116106  
 Qy 1418 TGTTCACCAAGGCTTGCTTTGAACCTCTGACTTCAGGTATTCACCCACTCAGCTCCCA 1477  
 Db 116107 TGTTCACCAAGGCTTGCTTTGAACCTCTGACTTCAGGTATTCACCCACTCAGCTCCCA 116166  
 Qy 1478 AAGCACTGGGATTACAGGATGAGCCAGCGTGCAGCCT 1517  
 Db 116167 AAGTCTGGGATTACAGGATGAGCCAGCGCTCCAGCCT 116206

## RESULT 17

ADQ97489

ID ADQ97489 standard; DNA; 48642 BP.

AC ADQ97489;

DT 07-OCT-2004 (first entry)

DE Human cancer associated sequence HD09-003, SEQ ID 466.

KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.

OS Homo sapiens.

PN WO2004060304-A2.

PD 22-JUL-2004.

PF 22-DEC-2003; 2003WO-US041389.

PR 27-DEC-2002; 2002US-00330773.

PA (SAGR-) SAGRES DISCOVERY INC.

PI Morris DW, Malandro MS;

DR WPI; 2004-543781/52.

PT New isolated cancer associated nucleic acids comprising at least 10 contiguous nucleotides, useful for diagnosing, preventing and/or treating cancers such as leukemia and lymphoma.

PS Claim 1; SEQ ID NO 466; 199pp; English.

XX The present invention relates to cancer associated sequences (ADQ97025-ADQ98004). The sequences are useful for the diagnosis, prevention and/or treatment of cancer, such as leukemia and lymphoma. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic formate directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 48642 BP; 11523 A; 11854 C; 12110 G; 13155 T; 0 U; 0 Other;

Query Match 10.3%; Score 205.6; DB 12; Length 48642;

Best Local Similarity 78.8%; Pred. No. 2,6e-41;

Matches 271; Conservative 0; Mismatches 69; Indels 4; Gaps 2;

Qy 1202 TATGTTCTCAGATCTTTTATTTTATTTTGAAGAGTCTCAGTTGTCAAC 1261  
 Db 8325 TAAAGGATATATCATCAGATTTTCTTTCTTTCTGAGACAGAGTCTCAGTTGTCAAC 8384  
 Qy 1262 CAGGCTGAGTACAGTGGCTGTGCTCGGCTCAGTGCACCTTGCCTCCAGGTTCA 1321  
 Db 8385 CAGGCTGAGTACAGTGGCTGTGCTCGGCTCAGTGCACCTTGCCTCCAGGTTCA 8444  
 Qy 1332 GCGATTTCCTG--TCAGTCTCCGATATGCTGGATATACAGGCGGACGC--ACACCAT 1377  
 Db 8445 GTGATTTCCTGCTCAGCTCCGAGTACGCTGGGATTAAGGCGCTGCAACATGCGCC 8504  
 Qy 1378 GCGTAAATTTTGTATTTTATTTAGTAGAGAGATTTGCGCATGTTGACAGGCTTGCCTTG 1437

Db 8505 GGGTACTTTTGTATTTTATTTAGTAGAGAGCGGGTTTCAACATGTTAGCCAGGCTGTTCCA 8564  
 Qy 1438 AACTCTTACCTTCAGGTATTCACCCACTCAGCTCCCAAGCATGAGATTACAGCA 1497  
 Db 8565 AACTCTTACCTTCAGGTATTCACCCACTCAGCTCCCAAGCATGAGATTACAGCA 8624  
 Qy 1498 TGAGCCACCGTGGCCAGCTGTTTCTCAGATCTGTAATTTGTT 1541  
 Db 8625 TGAGCCACCGAGCCGCGGACGACTTGACTTTTCTTTTCTTTT 8668

## RESULT 18

AAZ93815/C

ID AAZ93815 standard; DNA; 144460 BP.

AC AAZ93815;

DT 16-AUG-2000 (first entry)

DE Olfactory receptor operon.

KW Olfactory receptor protein; olfactory; smell; receptor; binding; detection; screening; genotyping; diallelic marker; human; ss.

OS Homo sapiens.

FH Location/Qualifiers

FT 2406..2600

FT /tag= a

FT /label= ORF1

FT 7521

FT /tag= m

FT /note= "Polymorphic base G or T"

FT 8192

FT /tag= n

FT /note= "Polymorphic base G or T"

FT 9711..10658

FT /tag= b

FT /label= ORF2

FT 14483

FT /tag= o

FT /note= "Polymorphic base A or T"

FT 19625

FT /tag= p

FT /note= "Polymorphic base C or T"

FT 20583

FT /tag= q

FT /note= "Polymorphic base A or G"

FT 24851..25369

FT /tag= c

FT /label= ORF3

FT 45714..46661

FT /tag= d

FT /label= ORF4

FT 76947

FT /tag= r

FT /note= "Polymorphic base A or C"

FT 80198..81115

FT /tag= e

FT /label= ORF 5

FT 91088

FT /tag= s

FT /note= "Polymorphic base G or C"

FT 91138

FT /tag= t

FT /note= "Polymorphic base A or G"

FT 91187

FT /tag= u

FT /note= "Polymorphic base A or G"

FT 96291..96902

FT /tag= f

FT /label= ORF6

FT 110758..111564

FT CDS

```
FT      /tag= g
FT      /label= ORP7
FT      complement(113644..114063)
FT      /tag= k
FT      /note= "Ubiquitin 1 pseudogene complement"
FT      CDS
FT      122525..122887
FT      /tag= h
FT      /label= ORP8
FT      complement(127489..127854)
FT      /tag= 1
FT      /note= "Ubiquitin 2 pseudogene complement"
FT      CDS
FT      132454..133889
FT      /tag= 1
FT      /label= ORP9
FT      allele
FT      133998
FT      /tag= v
FT      /note= "Polymorphic base A or G"
FT      allele
FT      140066
FT      /tag= w
FT      /note= "Polymorphic base C or T"
FT      allele
FT      141176
FT      /tag= x
FT      /note= "Polymorphic base A or G"
FT      CDS
FT      143398..143577
FT      /tag= j
FT      /label= ORP10
FT      allele
FT      144033
FT      /tag= Y
FT      /note= "Polymorphic base A or G"
FT      WO200021985-A2.
FT      PN
FT      20-APR-2000.
FT      PD
FT      XX
FT      13-OCT-1999; 99WO-1B001729.
FT      PR
FT      14-OCT-1998; 98US-0104299P.
FT      PA
FT      (GSEST ) GENSET.
FT      PI
FT      Bougueleret L, Malekzadeh K;
FT      DR
FT      WPI; 2000-317933/27.
FT      XX
FT      New nucleic acids encoding ten different olfactory receptor proteins and
FT      PT their allelic markers, are useful in genetic analysis and in screening
FT      PT for compounds which bind to the receptor proteins.
FT      PS
FT      Claim 1; Page 103-141; 155pp; English.
FT      CC
FT      Ten new olfactory receptor proteins and their allelic markers have been
FT      CC described. The sequences encoding these receptor proteins and which
FT      CC contain the allelic markers can be used for genotyping. The olfactory
FT      CC receptor proteins can be used to screen for substances which bind to
FT      CC them. See GENSEQ records AA293816-25 and AA783386-95
FT      CC
FT      SQ
FT      Sequence 144460 BP; 46068 A; 27088 C; 26615 G; 44676 T; 0 U; 13 Other;
FT      Query Match 10.2%; Score 204.8; DB 3; Length 144460;
FT      Best Local Similarity 63.8%; Pred. No. 6.5e-41;
FT      Matches 394; Conservative 0; Mismatches 212; Indels 12; Gaps 5;
QY      1155 CATTTCAGGTGGAATTGTCAGAGAAATATTGATTGATTGTTTCTCAGATCTTTTAT 1224
DB      57682 CTTTGGAGAGTGTATATGTCCTTGGCCACTTTATATATGATTTTATTTATTTA 57623
QY      1225 TTTTATTTTGAACAGAGTCTCACTTGTACCCAGGCTGAGTACAGTGGCTTG 1284
DB      57622 TTTTATTTATTTAGACAGATCTCACTGTGCGCCAGGCTGAGTGAAGTGG-TGCA 57564
QY      1285 GTCTGGGCTCACTGCAACTCTGCTCCAGGTTCAAGGATTTCTTG-TTCAAGTTCC 1342
DB      57563 ATCTGGGCTCACTGCAACTCTCACTCTGAGATTCAAGGATTTCTCAAGCTCAGCTCT 57504
```

```
QY      1343 CGAATAGCTGGAGATTACAGGCGCATGCAACCATGCC---TAATTTTGATTTTAG 1398
DB      57503 AGAGTAGCTGGAGATTACAG--GCACGACCAACCATGCCAGTATTTTGTATTTTAG 57446
QY      1399 TAGAGACAGAGTTTCGCGCATGTTTACACAGGCTTGCCTTGAACCTGACCTCAGGTATC 1458
DB      57445 TAGAACAAGGTTTGGCATGTTGCGCAGGCTGATCTTGAACCTGACCTCAAGATTAATTC 57386
QY      1459 CACCACTCAGCTCCCAAGACACTGGGATTTACAGGATGAGCACCGCCGAGCCTG 1518
DB      57385 CACCTGCTCTGCTCTCCCAAGTGTGCGATTTACTGGATGAGCAGCTGTGGCAGCATC 57326
QY      1519 TTTTCTCAGATCC---TGATTTGTTTCTGAAGCCTTCATTTCTTATTCATTT 1575
DB      57325 TTAGCCCACTTTTAATGTTGTTGTTGCTTTCTTGTAAATTTAAGTTCTTAAACAT 57266
QY      1576 TGGAAATGATCACTCAAGTAAAGTTTAAACATCAATCTTTGGAATCCCTGG 1635
DB      57265 GCTGAATATTTAGACCTTTTGTCAATACATTAAGTTTGGAAATATTTCTTCTATTCATAGG 57206
QY      1636 TTCTTTCTTATTCCTACAAAATATATGTTCAAGTATAGCTGATGTTTCTTTCAAT 1695
DB      57205 CTGCTTTTACTCTGTTACAGATTGTTTGTGCTGTAAGAAGCTCTTAAGTTTAATTAG 57146
QY      1696 TATTCATTTCTCTATCTCAGAAATTTATCTCARGCCTAATGTTATGATGCTTCACT 1755
DB      57145 ATCTACTTGTCAATTTTGTCTTTTCTTTTGTTCGAATTTGCTTTGTCTTTGATG 57086
QY      1756 TCTTGTCAATCCAGTTTCT 1773
DB      57085 AATTCATTTGCCCATTTCT 57068
RESULT 19
ACN44262
ID ACN44262 standard; DNA, 168821 BP.
XX
ACN44262;
XX
AC
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG18035.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PR 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 622; Opp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
```



CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published  
XX  
SQ Sequence 168821 BP; 39588 A; 43389 C; 45655 G; 40189 T; 0 U; 0 Other;  
Query Match 10.2%; Score 204.8; DB 11; Length 168821;  
Best Local Similarity 77.7%; Pred. No. 6,9e-41;  
Matches 286; Conservative 0; Mismatches 77; Indels 5; Gaps 3;  
OY 1191 AAGTATTTGATTAGTTTCTGATCTTTTATTTTATTTTATTTTGGAAACAGAGTCTC 1250  
DB 166269 AATCCTATATTTATTTATTTATTTATTTATTTATTTATTTATTTGAGACGGTGTCTT 166328  
OY 1251 ACTTGTGACCCGAGGCTGAGTACAGTGGCTGGCTGCTGCTGCTGCTGCTGCTGCT 1310  
DB 166329 GCTTATACCCAGGCTGAGTACAGTGGC-ACAATCTAGCTCAGTCAACCTCCACT 166387  
OY 1311 CCCAGGTTCAAGCGATTCTCTG--TCAGCTTCCGAAATAGCTGGGATTACAGCGCATG 1368  
DB 166388 CCCAGGTTCAAGCGATTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 166447  
OY 1369 C-ACCACCAGCCTAATTTTGTATTTTGTATTTTGTATTTGTATTTGTATTTGTATTTGAC 1426  
DB 166448 CCACCAACGCGCGTAAATTTTGTATTTTGTATTTGTATTTGTATTTGTATTTGTATTTGAG 166507  
OY 1427 GCGTTCCTGTAACCTCGATCTGATGATCCACCCACCTGCTGCCAAGACATGG 1486  
DB 166508 GCGTTCCTGTAACCTCGATCTGATGATCCACCCACCTGCTGCCAAGACATGG 166567  
OY 1487 GATTACAGGATGAGCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1546  
DB 166568 GATTACAGGATGAGCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 166627  
OY 1547 AGCCTTCA 1554  
DB 166628 AATCCTCA 166635  
RESULT 20  
ABAI8360  
ID ABA18360 standard; DNA; 11343 BP.  
XX AC  
XX ABA18360;  
XX  
XX 23-JAN-2002 (first entry)  
DE Human nervous system related polynucleotide SEQ ID NO 10691.  
XX  
XX Human; noctropic; neuroprotective; cytosolic; dermatological; virucide;  
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnereary;  
XX antiparkinsonian; antisticking; antianaemic; antidiabetic; cancer;  
XX antitubercular; hepatotropic; cerebroprotective; antiinflammatory;  
XX antiallergic; antidiabetic; antidiabetic; antidiabetic; antifungal;  
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
XX  
XX 16-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001334.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 18-APR-2000; 2000US-0190075P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214866P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227099P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 06-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231245P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.



KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;  
KM gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; db.  
XX  
OS Homo sapiens.  
XX  
PN WO200292787-A2.  
XX  
PD 21-NOV-2002.  
XX  
PF 26-MAR-2002; 2002WO-US009257.  
XX  
XX 27-MAR-2001; 2001US-0278650P.  
PR 12-SEP-2001; 2001US-00950082.  
PR 12-SEP-2001; 2001US-00950083.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR MPI; 2003-129287/12.  
XX  
PT New human secreted proteins and nucleic acid molecules, useful for  
PT preparing a diagnostic or pharmaceutical composition for diagnosis,  
PT preventing or treating hematopoietic or hematologic disorders, e.g.  
PT anemia or hemophilia.  
XX  
PS Disclosure; SEQ ID NO 983; 1512bp; English.  
XX  
XX The invention comprises the amino acid and coding sequences of human  
CC secreted proteins. The DNA and protein sequences of the invention are  
CC useful for detecting, preventing, diagnosing, prognosticating, treating  
CC or ameliorating: hematopoietic or hematological disorders (e.g. anaemia  
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease  
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);  
CC wound healing and disorders of epithelial cell proliferation; immune  
CC disorders (e.g. autoimmune disorders and asthmatic disorders);  
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);  
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);  
CC and gastrointestinal disorders (e.g. duodenal ulcers and  
CC gastroenteritis). The present DNA sequence was used in the  
CC exemplification of the invention.  
XX  
SQ Sequence 11343 BP; 3085 A; 2464 C; 2534 G; 3260 T; 0 U; 0 Other;  
Query Match 10.2%; Score 203.4; DB 10; Length 11343;  
Best Local Similarity 70.0%; Pred. No. 5.2e-41;  
Matches 317; Conservative 0; Mismatches 131; Indels 5; Gaps 3;  
QY 1125 TATTAAGAGATTTTATGAGCATGTGTGACAAAGAGAGGCAATTCAGTGTGAATTGT 1184  
DB 2027 TTTAAGAGCCCTGTGAACATCTGTGAACATTAACATTCCTTTGAGGCTGTG 2086  
QY 1185 TCAGAGAAATTTGATTAATGTTTCTCAGATCTTTTATTTTATTTTGAACAG 1244  
DB 2087 TTAGATTAAGTAGAGTTTGTCTTAATTAATGTTGGTTTGTGTTGTTGAGACG 2146  
QY 1245 AGTCTCACTTGTACCCAGGCTGAGTACAGTGTGTGTCTGCTCCTACTGCAACT 1304  
DB 2147 ACTCTCACTGTGTGCGCCAGGCTGAGTGTGAGTGTGC-GCATTCTTGATCCTGCAACT 2205  
QY 1305 CTGCTCCAGGTTCAAGGATTTCTCTG--TCAGTTTCCGAATGCTGGGATTAACAG 1362  
DB 2206 CTGCTCCAGGTTCAAGGATTTCTCTGAGCTCTCCTGAGTGTGAGGATTAACAG 2265  
QY 1363 CCATGCAACCATC--CTAATTTTGTATTTTATAGAGACAGATTTGCGCATGT 1420  
DB 2266 TGATTCACACGAGCCAGCTAATTTTGTATTTTATAGAGAGTGGGTTTCAACATGT 2325  
QY 1421 TCAGCAGGCTTGCTTGAACCTCTGACTTGAGTATCAACCCACTCAGCTCCCAAG 1480  
DB 2326 TGCCAGGCTGTCTGAACTCTCTGACCTCAAGTATCCACCGGCTCCTCAAG 2385  
QY 1481 CACTGGATTAAGAGATAGAGCAGCGTCCGAGCTGTTTCTCAGATCTGTATTTGT 1540

DB 2386 GGCTGGATTAAGAGTGTGAGGCACTGTGCCAGGCAAACTTTTATAGACAGA 2445  
QY 1541 TTCGAGCCTTCATTTCTATCTTCTATTCAT 1573  
DB 2446 ATTATACCTCACAGATCATCATTTTCTT 2478  
RESULT 22  
ABT17025  
ID ABT17025 standard; DNA, 11343 BP.  
XX  
AC ABT17025;  
XX  
DT 03-APR-2003 (first entry)  
XX  
XX Human secreted protein-related DNA sequence - SEQ ID NO 379.  
DE  
XX  
XX Human; gene; ds; protein therapy; immediate hypersensitivity disease;  
KM allergic disorder; asthmatic disorder; gene therapy; secreted protein;  
KM hay fever; allergic conjunctivitis; allergic rhinitis;  
KM binding partner identification; chromosome identification;  
KM radiation hybrid mapping; long-range restriction mapping.  
XX  
OS Homo sapiens.  
XX  
PN WO200277186-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 26-MAR-2002; 2002WO-US009239.  
XX  
PR 27-MAR-2001; 2001US-0278650P.  
PR 12-SEP-2001; 2001US-00950082.  
PR 12-SEP-2001; 2001US-00950083.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR MPI; 2003-175010/17.  
XX  
PT Use of human secreted proteins and nucleic acids for preparing a  
PT diagnostic or pharmaceutical composition for diagnosing or treating  
PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic  
PT conjunctivitis or rhinitis.  
XX  
PS Disclosure; Page 820-823; 823bp; English.  
XX  
XX The invention comprises the amino acid and coding sequences of human  
CC secreted proteins. The DNA and protein sequences of the invention are  
CC useful for the diagnosis and treatment of allergic disorders, asthmatic  
CC disorders and immediate hypersensitivity diseases (e.g. hay fever,  
CC allergic conjunctivitis and allergic rhinitis). The proteins of the  
CC invention are also useful for identifying a binding partner. The nucleic  
CC acids of the invention are also useful for chromosome identification,  
CC radiation hybrid mapping or long-range restriction mapping. The present  
CC DNA sequence represents a human secreted protein-related DNA sequence  
SQ  
Sequence 11343 BP; 3085 A; 2464 C; 2534 G; 3260 T; 0 U; 0 Other;  
Query Match 10.2%; Score 203.4; DB 10; Length 11343;  
Best Local Similarity 70.0%; Pred. No. 5.2e-41;  
Matches 317; Conservative 0; Mismatches 131; Indels 5; Gaps 3;  
QY 1125 TATTAAGAGATTTTATGAGCATGTGTGACAAAGAGAGGCAATTCAGTGTGAATTGT 1184  
DB 2027 TTTAAGAGCCCTGTGAACATCTGTGAACATTAACATTCCTTTGAGGCTGTG 2086  
QY 1185 TCAGAGAAATTTGATTAATGTTTCTCAGATCTTTTATTTTATTTTGAACAG 1244  
DB 2087 TTAGATTAAGTAGAGTTTGTCTTAATTAATGTTGGTTTGTGTTGTTGAGACG 2146

Oy	1245	AGTCGACTTTGTGTAACCCAGGCTGGAGATCAAGTGGCTGAGTCTCGCTCACTGCAACT	1304
Db	2147	AGTTCACCTCTGTGCGCCAGGCTGAGTGCAAGTGCC-CCATCTTGGATCACTGCAACT	2205
Oy	1305	CTGCCTCCAGGATTCAAGCAGATTCTCCGCT--TCAGCTCCCGAATAGCTGGAAATTACAG	1362
Db	2206	CTGCCTCCAGGATTCAAGCGAGTCTCTGCTCAGCTCTCCTAGTAGCTGGAAATTACAG	2265
Oy	1363	CGCATGCACCAACCAAGC--CTAATTTTGTATTTTATAGAGACAGAGTTTCGCATGT	1420
Db	2266	TGCATACCAACCAAGCCGCAAGTAAATTTTGTATTTTATAGAGAGTGGGTTTCAACAGT	2325
Oy	1421	TGACCAAGCTTGCTTGAATCTCTGACTTCAGTGTATCCACCACCTCAGCTTCCCAAG	1480
Db	2326	TGGCAGAGCTGGTCTCGAATCTCTGACTCCTCAAGTATCAACCGGCTCAGCTCTTAAG	2385
Oy	1481	CACATGGGATTACAGGATGAGGCCACCGTGCAGCTGTTTTCTAGATCCTGATTTTGT	1540
Db	2386	GAGCTGGGATTACAGGTGTGAGGCCATGTGCCACGCAAGGCAAACTTTTTTTTAGCAGA	2445
Oy	1541	TTCTGAAGCCTTCAATTCATCTCTTAATCAT	1573
Db	2446	ATTATACCTCACAGATCATATTTTTTTCTT	2478

RESULT 23  
ABZ68170  
ID ABZ68170 standard; DNA; 11343 bp.

AC	ABZ68170;
XX	
DT	26-MAR-2003 (first entry)

Human secreted protein encoding genomic DNA SEQ ID NO 1693

KM Human; secreted protein; nootropic; neuroprotective; cytosolic;  
KM viruslike; dermatological; immunosuppressive; anti-HIV;  
KM vulnery; antibacterial; antiparkinsonian; antiskicking; antanaemic;  
KM antiaftrictic; cancer; antitrematic; hepatotropic; cerebroprotective;  
KM antiflamatory; antiallergic; antidiabetic; antitumor; anticonvulsant;  
KM antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;  
KM cardiovascular disorder; neurological disease; nephrotropic;  
KM gene therapy; gene; ds.

OS Homo sapiens.

PN WO200277186-A2.

PD 03-OCT-2002.

PF 26-MAR-2002; 2002WO-US009188.

PR 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950083.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2003-040583/03.

PT New human secreted p

PT multiple sclerosis,

XX

XX

CC encoded secreted pro

CC therapy. The genes a

CC	in the specification. The nucleic acids, proteins, antibodies and
CC	(ant) agonists are useful in the diagnosis, treatment and prevention of:
CC	(a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC	adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC	lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC	allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC	diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC	arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC	myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC	cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC	bacterial, fungal and parasitic infections
XX	
SQ	Sequence 11343 BP; 3085 A; 2464 C; 2534 G; 3260 T; 0 U; 0 Other;
Query Match	10.2%; Score 203.4; DB 10; Length 11343;
Best Local Similarity	70.0%; Pred. No. 5.2e-41;
Matches 317; Conservative	0; Mismatches 131; Indels 5; Gaps 3;
Db	1125 TATAAGAGATTTTTCATGACATGTGTACAAGAGGGCGCATTTTCAGTGTTGAATTGT 1184
2027	TTTAAAGACCCTGGTAACATCTGTTAACAATTAAGTATCAATTTGCCITTCAGGACTGG 2086
OY	1185 TCAGAAGAATTTTGCATTAATGTTTTCTCAGATCTTTTATTTTATTTTGGAAACAG 1244
Db	2087 TTAGATATACGTAGTGTGTTTCTTAATTAATGTTGGGTTTTTTTGTGTTGGAGCAG 2146
OY	1245 AGTCTCATTTTGTACCCAGCGCTGGAGTACATGTGCTGTGCTCGCTCACATGCCA 1304
Db	2147 AGTCTCATCTGTGTGCCAGCGCTGGAGTGGACGTGGC-GCATCTTGTGATCACATCCA 2205
OY	1305 CTGCTCTCCAGAGTTCAGACGATTCCTCG--TCAGCTTCCCGAATAGCTGGGATTCAGG 1362
Db	2206 CTGCTCTCCAGAGTTCAGACGAGTCTCTGCTCGCTCGCTCTCTTAGTACTGGGATTCAGG 2265
OY	1363 CGCATGCACACACATGC-CCTAATTTTTGATTTTTTAGTAGACAGAGTTTCGCCATGT 1420
Db	2266 TGCAATACACACAGCGCCAGCATTAATTTTTTGATTTTTTAGTAGAGGTGGGTTTCACCATGT 2325
OY	1421 TGACCAAGCTTGCTTTGAACTCTGACTTCAGGTGATTCACACCACTCAGCTCCCAAAG 1480
Db	2326 TGCCAGAGCTGTCTCGAACCTCTGACCTCAAGTATCCACCCGCTCAGCCTCCCTAAG 2385
OY	1481 CACGTGGATTCAGAGCATGAGCCACCGGCCAGCCTGTTTTCTCAGATCCTGATTTGT 1540
Db	2386 GGCTGGGATTCAGAGTGTGAGCCACTGTGCCACGCGGCAAACTTTTTTTTAGCAGGA 2445
OY	1541 TTCGAAGCTTCATTTCTATCTTTTTCAT 1573
Db	2446 ATTATACCTCACACGATCATCTTTTTCCTT 2478
RESULT 24	
ACN44586/C	
ID	ACN44586 standard; DNA; 118931 BP.
AC	ACN44586;
XX	
DT	18-NOV-2004 (first entry)
XX	
DE	Human genomic sequence hCG30014.
XX	
KW	Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
OS	Homo sapiens.
XX	
PX	MO2003073826-AZ.
PD	
PF	12-SEP-2003.
XX	
PR	28-FEB-2003; 2003WO-US006235.
XX	
XX	01-MAR-2002; 2002US-00087192.
XX	

PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW;  
XX WPI; 2003-328604/31.  
XX  
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
PT comprises a nucleotide sequence.  
XX  
PS Claim 1; SEQ ID NO 1108; Opp: English.  
XX  
XX The present invention relates to novel DNA and protein sequences which  
CC are associated with carcinomas. The sequences are useful for: (i) for  
CC screening drug candidates; (ii) for screening of bioactive agent capable  
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biobip;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published  
XX  
SQ Sequence 118931 BP; 33837 A; 24131 C; 23530 G; 35376 T; 0 U; 2057 Other;  
XX  
Query Match 10.1%; Score 202.8; DB 11; Length 118931;  
Best Local Similarity 83.6%; Pred. No. 1.9e-40;  
Matches 266; Conservative 0; Mismatches 47; Indels 5; Gaps 3;  
QY 1216 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTACCCAGGCTGAGATA 1275  
DB 76898 TCTTTTCTTTTATTTTATTTTGAACAGAGTCTCACTTGTGCCCCAGGCTAGAGTACA 76839  
QY 1276 GTGGCTGTGCTCGGCTCACTGCAACCTCTGCTCCGAGTTCAAGGATTCTCCTG-- 1333  
DB 76838 GTGG-TGCCATCTCGGCTCACTGCAACCTCTGCTCCAGGTCACAGCCATTCTCCTGCC 76780  
QY 1334 TCAGCTTCCCGAATAGCTGGGATTAACAGCGCATGACACCA--TGCCATAATTTTGTGA 1391  
DB 76779 TCAGCTTCCCGAGTAGCTGGGATTAACAGCGCATGACACCACTGGGTAATTTCTGTA 76720  
QY 1392 TTTTATAGTAGACAGAGTTTGGCCATGTTACACAGGCTTGCCTTGAATCTCTGACTTCA 1451  
DB 76719 TTTTATAGTAGACAGAGGTTTACCATGTTGGCAGGCTGCTCTGAACTCTGACTTCA 76660  
QY 1452 GGTGATCCACCCAGCTCAGCTCCCAAGACATGGGATTAACAGGATGAGCCACCGTGCC 1511  
DB 76659 GGTGATCCACCTGTCTGCGCTTCCCAAGTGTGGGATTAACAGGATGAGCCACCGTGCC 76600  
QY 1512 CAGCGTGTCTTCAGAT 1529  
DB 76599 CGGCCCGCTGCACTCCTAT 76582  
RESULT 25  
ADH19244  
ID ADH19244 standard; DNA; 124164 BP.  
XX  
AC ADH19244;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Human 5-hydroxytryptamine receptor type 3B SNP variant DNA.  
XX  
XX HTR3B; 5-hydroxytryptamine receptor type 3B; 5-HT; antiemetic;  
KW trianquiliser; neuroleptic; antialcoholic; antimigraine; analgesic;  
KW gastroenteric; setrone; central nervous system; drug treatment;  
KW postoperative nausea; vomiting; chronic fatigue syndrome;  
KW obsessive-compulsive disorder; schizophrenia; alcoholism; anxiety;  
KW tourette syndrome; migraine; headache; gastrointestinal motility;

KW cancer chemotherapy; forensic marker; ds; human; SNP;  
KW single nucleotide polymorphism; gene.  
XX  
OS Homo sapiens.  
XX  
PN WO2003097873-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-EP005120.  
XX  
PR 15-MAY-2002; 2002EP-00010209.  
XX  
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.  
XX  
PI Brockmoeiller HJ;  
XX  
DR WPI; 2004-022892/02.  
XX  
PT New 5-Hydroxytryptamine receptor type 3B polynucleotide, useful for  
PT diagnosing and/or treating a setrone-treatable disease such as disorders  
PT of central and/or peripheral nervous system e.g., schizophrenia.  
XX  
PS Claim 1; Page; 150pp; English.  
XX  
XX The invention relates to a novel polynucleotide encoding an HTR3B (5-  
CC hydroxytryptamine [5-HT] receptor type 3B) polypeptide or fragment having  
CC an amino acid substitution. The polynucleotide of the invention  
CC demonstrates antiemetic, trianquiliser, neuroleptic, antialcoholic,  
CC antimigraine, analgesic and gastrointestinal activities and may be useful  
CC in preparing a composition for diagnosing or treating a disease,  
CC particularly a setrone-treatable disease. Such a disease or dysregulation  
CC is related to the central and peripheral nervous system or secondary to  
CC drug treatment, such as postoperative nausea and/or vomiting, chronic  
CC fatigue syndrome, obsessive-compulsive disorders, schizophrenia,  
CC alcoholism, anxiety disorders, tourette syndrome, migraine, headache and  
CC gastrointestinal motility disorders, preferably nausea and/or vomiting  
CC secondary to cancer chemotherapy. The polynucleotides and polypeptides  
CC may also be useful as forensic markers. The current sequence is that of  
CC the human HTR3B SNP variant DNA of the invention in which bases 36678-  
CC 36680 of the alternative SNP form have been deleted. This sequence is not  
CC shown in the specification per se but was created by the indexer using  
XX information from claim 1.  
SQ Sequence 124164 BP; 32694 A; 28609 C; 27969 G; 34892 T; 0 U; 0 Other;  
XX  
Query Match 10.1%; Score 202.6; DB 12; Length 124164;  
Best Local Similarity 84.3%; Pred. No. 2.2e-40;  
Matches 264; Conservative 0; Mismatches 44; Indels 5; Gaps 3;  
QY 1216 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTACCCAGGCTGAGATA 1275  
DB 16571 TTTTATTTTATTTTATTTTGAAGAGGAGTCTCACTTGTGACCCAGGCTGAGTCA 16630  
QY 1276 GTGGCTGTGCTCGGCTCACTGCAACCTCTGCTCCGAGTTCAAGGATTCTCCTG-- 1333  
DB 16631 GTGG-TGCCATCTCGGCTCACTGCAACCTCTGCTCTGTTCAAGGATTCTCCTGCC 16689  
QY 1334 TCAGCTTCCCGAATAGCTGGGATTAACAGCGCATGAC--ACCAACATGCTAATTTTGTGA 1391  
DB 16690 TCAGCTTCTGATAGTACCTGGGATTAACAGTGTGCTGCCACCACTGCTGATTTTGTGA 16749  
QY 1392 TTTTATAGTAGACAGAGTTTGGCCATGTTACACAGGCTTGCCTTGAATCTCTGACTTCA 1451  
DB 16750 TTTTATAGTAGACAGGAGTTTACCATGTTGGCCAGGCTGAGTGAATCTCCGACCTCA 16809  
QY 1452 GGTGATCCACCCAGCTCAGCTCCCAAGACATGGGATTAACAGGATGAGCCACCGTGCC 1511  
DB 16810 GGTGATCCACCACTGTCTCTCCCAAGTGTGGGATTAACAGGATGAGCCACCGTGCC 16869  
QY 1512 CAGCGTGTCTTCT 1524  
DB 16870 CAGCGTGTACCT 16882

ID	ADH19246	standard; DNA; 124165 BP.
AC	ADH19246;	
XX	11-MAR-2004	(first entry)
DE	Human 5-hydroxytryptamine receptor type 3B SNP variant DNA 2.	
KW	HTT3B; 5-hydroxytryptamine receptor type 3B; 5-HT; antiemetic;	
KW	tranquilliser; neuroleptic; antialcoholic; antimigraine; analgesic;	
KW	gastrointestinal; serotonin; central nervous system; drug treatment;	
KW	postoperative nausea; vomiting; chronic fatigue syndrome;	
KW	obsessive-compulsive disorder; schizophrenia; alcoholism; anxiety;	
KW	tourrette syndrome; migraine; headache; gastrointestinal motility;	
KW	cancer chemotherapy; forensic marker; ds; human; SNP;	
XX	single nucleotide polymorphism; gene.	
OS	Homo sapiens.	
PN	MO2003097873-A2.	
PD	27-NOV-2003.	
PF	15-MAY-2003; 2003MO-EP005120.	
PR	15-MAY-2002; 2002EP-00010209.	
XX	(EPID-) EPIDAUROS BIOTECHNOLOGIE AG.	
XX	Brockmeoeller HJ;	
XX	WPI; 2004-022892/02.	
DR	New 5-Hydroxytryptamine receptor type 3B polynucleotide, useful for	
PT	diagnosing and/or treating a serotonin-treatable disease such as disorders	
PT	of central and/or peripheral nervous system e.g., schizophrenia.	
XX	Claim 1; Page; 150pp; English.	
XX	The invention relates to a novel polynucleotide encoding an HTT3B (5-	
CC	hydroxytryptamine [5-HT] receptor type 3B) polypeptide or fragment having	
CC	an amino acid substitution. The polynucleotide of the invention	
CC	demonstrates antiemetic, tranquilliser, neuroleptic, antialcoholic,	
CC	antimigraine, analgesic and gastrointestinal activities and may be useful	
CC	in preparing a composition for diagnosing or treating a disease,	
CC	particularly a serotonin-treatable disease. Such a disease or dysregulation	
CC	is related to the central and peripheral nervous system or secondary to	
CC	drug treatment, such as postoperative nausea and/or vomiting, chronic	
CC	fatigue syndrome, obsessive-compulsive disorders, schizophrenia,	
CC	alcoholism, anxiety disorders, tourrette syndrome, migraine, headache and	
CC	gastrointestinal motility disorders, preferably nausea and/or vomiting	
CC	secondary to cancer chemotherapy. The polynucleotides and polypeptides	
CC	may also be useful as forensic markers. The current sequence is that of	
CC	the human HTT3B SNP variant DNA 2 of the invention in which bases 64500-	
CC	64501 of the alternative SNP form have been deleted. This sequence is not	
CC	shown in the specification per se but was created by the indexer using	
CC	information from claim 1.	
XX	Sequence 124165 BP; 32695 A; 28608 C; 27970 G; 34892 T; 0 U; 0 Other;	
XX	Query Match 10.1%; Score 202.6; DB 12; Length 124165;	
XX	Best Local Similarity 84.3%; Pred. NO. 2.2e-40;	
XX	Matches 264; Conservative 0; Mismatches 44; Indels 5; Gaps 3;	
XX	1216 TCTTTTATTTTATTTTATTTTATTTTGAACAGATCTCACTTTGTCAACCAAGGCTGAGTACA 1275	
XX	16571 TTTTATTTTATTTTATTTTATTTTATTTTGTGAGAGAGATCTCACTTTGTGACCAAGGCTGAGATGCA 16630	
XX	1276 GTGGCTGTGTCCTGAGCTCACTGCAACCTTGCCTCCACAGATTCAACGATTTCTCTG-- 1333	

Db	16631	GTGG-TGGCATCTCACTCAACCAACACCTCTGCTCTCTGTTTCAAGCATTCCTCTGCC	16689
Qy	1334	TCAGCTTCCGGAATACCTGGGATTTACAGCGCATGC--ACACCATGCTTAATTTTGTGA	1391
Db	16690	TCAGCCTCTCGAGTAAGTGGGATTTACAGGTGCCTGCCACATACCGCTGGCTAATTTTGTGA	16749
Qy	1392	TTTTTGTATGAGACAGAGTTTGGCCATGTTGACACAGGCTTGCCTTGAATCCTGACTTCA	1451
Db	16750	TTTTTGTATGAGACAGCGGGTTTCCACAGTTTGGCCAGGCTGGGGTTGAATCTCCGACCTCA	16809
Qy	1452	GGTATGATCAACCCACCTCAGGCTCCCAACACACTGGGATTTACAGGATGAGCCACCGTCC	1511
Db	16810	GGTATGATCAACCCACCTCTGCTCTCCCAAGTGTGGATTTACAGGACATGAGCCACTGCCCC	16859
Qy	1512	CAGCCTGTTTTCT	1524
Db	16870	CAGCCTGTACT	16882
RESULT 27			
ID	ADP83402		
AC	ADP83402	standard; DNA; 124167 BP.	
XX	ADP83402;		
DT	26-FEB-2004	(first entry)	
DE	Human 5-hydroxytryptamine receptor type 3 gene.		
XX	Human; antiemetic; setrone; 5-hydroxytryptamine receptor type 3;		
KM	receptor; single nucleotide polymorphism; SN; HTR3B; gene; ds.		
OS	Homo sapiens.		
XX	WO2003100091-A1.		
PN	04-DEC-2003.		
PD	22-MAY-2003; 2003WO-EP005366.		
PF	24-MAY-2002; 2002EP-00011491.		
XX	(EPID-) EPIDUROS BIOTECHNOLOGIE AG.		
PA	Brockmøller HJ;		
PI	WPI; 2004-035165/03.		
XX	GENBANK; AF001874.		
DR			
XX			
PT	Use of setrones for preparing a pharmaceutical composition for treating		
PT	or preventing setrone-treatable diseases in a subject having in its		
PT	genome less than three copies of a polynucleotide encoding a functional		
PT	CYP2D6 polypeptide.		
XX			
PS	Disclosure; SEQ ID NO 52; 153bp; English.		
CC	The present sequence comprises the human 5-hydroxytryptamine receptor		
CC	type 3 HTR3B gene. In a variant of the gene ADP83375, nucleotides 36678-		
CC	36680 (AAG) of this sequence are deleted. The invention relates to the		
CC	use of setrones (antiemetics) for treating and/or preventing setrone-		
CC	treatable diseases in a subject having in its genome fewer than 3 copies		
CC	of a polynucleotide encoding a functional CYP2D6 polypeptide, and also		
CC	having in its genome a second variant allele comprising a polynucleotide		
CC	capable of hybridising to a HTR3B gene having ACG at positions 36678-		
CC	36680. The treatment regimen can be modified according to the genotype of		
CC	the subject's CYP2D6 and/or HTR3B gene. Non-responders to antiemetic		
CC	therapy can be identified on a pharmacogenetic basis, allowing a suitable		
CC	therapy to be selected. The setrone-treatable diseases are postoperative		
CC	nausea and/or vomiting, or nausea and/or vomiting secondary to cancer		
CC	chemotherapy, radiation therapy, migraine, acetaminophen poisoning,		
CC	prostaglandin therapy, and opioid treatment, spinal or epidural opioid-		
CC	related pruritus, acute levodopa-induced psychosis, bulimia nervosa,		

CC fibromyalgia, chronic fatigue syndrome, obsessive-compulsive disorders,  
 CC schizophrenia, alcoholism, cocaine addiction, opioid withdrawal syndrome,  
 CC drug withdrawal phenomena, anxiety disorders, cognitive disturbance,  
 CC neuroleptic-induced tardive dyskinesia, Tourette's syndrome, migraine  
 CC headache or gastrointestinal motility disorder (all claimed).

XX Sequence 124167 BP; 32696 A; 28609 C; 27970 G; 34892 T; 0 U; 0 Other;

Query Match 10.1%; Score 202.6; DB 12; Length 124167;

Best Local Similarity 84.3%; Pred. No. 2.2e-40; Matches 264; Conservative 0; Mismatches 44; Indels 5; Gaps 3;

QY 1216 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTACCCAGGCTGAGTACA 1275  
 DB 16571 TTTTTTTTTTTTTTTTTTTTGTAGAGGAGTCTCACTTTGTAGCCAGCTGGAGTGA 16630  
 QY 1276 GTGGCTGTGTCTCGGCTACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG-- 1333  
 DB 16631 GTGG-TGGCATCTCAGCTACCAACAACCTCTGCTCTGTGTTCAAGGATTTCTCTGCC 16689  
 QY 1334 TCAGCTTCCGAATAGCTGGGATTAAGGGGATG--ACACACATGGCTAATTTTGTGA 1391  
 DB 16690 TAGGCTCTTGAGTAACTGGGATTAAGGATGCTGACACATACCTGGCTAATTTTGTGA 16749  
 QY 1392 TTTTATAGTAGACAGAGTTTGGCCATGTTGACAGGCTTGGCTTGAATCTCTGACTTCA 1451  
 DB 16750 TTTTATAGTAGACAGGAGGTTTACCATGTTGGCCAGGCTGGGTTGAATCTCCGACTCA 16809  
 QY 1452 GGTGATCCACCACTCAGCTCCCAAGACACTGGGATTAAGGATGAGCAACCGTGGC 1511  
 DB 16810 GGTGATCCACCACTCAGCTCCCAAGGATGAGGATTAAGGATGAGCAACCGTGGC 16869  
 QY 1512 CAGCTGTTTCT 1524  
 DB 16870 CAGCTGTTTCT 16882

RESULT 28

ADH19240

ID ADH19240 standard; DNA, 124167 BP.

AC ADH19240;

DT 11-MAR-2004 (first entry)

XX Human 5-hydroxytryptamine receptor type 3B gDNA.

KM HTR3B; 5-hydroxytryptamine receptor type 3B; 5-HT; antiemetic;  
 KM trianquilliser; neuroleptic; antialcoholic; antimigraine; analgesic;  
 KM gastrointestinal; setrone; central nervous system; drug treatment;  
 KM postoperative nausea; vomiting; chronic fatigue syndrome;  
 KM obsessive-compulsive disorder; schizophrenia; alcoholism; anxiety;  
 KM tourette syndrome; migraine; headache; gastrointestinal motility;  
 KM cancer chemotherapy; forensic marker; ds; human; SNP;  
 KM single nucleotide polymorphism; gene.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH variation replace(36777,A)

FT variation /tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(63724,T)

FT variation /tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(63725,G)

FT variation /tag= c  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(64152,C)

FT variation /tag= d  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(64228,A)

FT variation /tag= e

FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(64757,T)  
 FT /tag= f  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(64910,A)  
 FT /tag= g  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(65011,A)  
 FT /tag= h  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(74691,T)  
 FT /tag= i  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(74737,A)  
 FT /tag= j  
 FT /standard\_name= "Single nucleotide polymorphism"

PN WO2003097873-A2.

PD 27-NOV-2003.

PF 15-MAY-2003; 2003WO-EP005120.

PR 15-MAY-2002; 2002EP-00010209.

PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

PI Brockmoeiller HJ;

DR WPI; 2004-022892/02.

DR P-PSDB; ADH19241, ADH19243.

PT New 5-Hydroxytryptamine receptor type 3B polynucleotide, useful for  
 PT diagnosing and/or treating a setrone-treatable disease such as disorders  
 PT of central and/or peripheral nervous system e.g., schizophrenia.

XX Claim 1; SEQ ID NO 49; 150bp; English.

XX The invention relates to a novel polynucleotide encoding an HTR3B (5-  
 CC hydroxytryptamine [5-HT] receptor type 3B) polypeptide or fragment having  
 CC an amino acid substitution. The polynucleotide of the invention  
 CC demonstrates antiemetic, tranquiliser, neuroleptic, antialcoholic,  
 CC antimigraine, analgesic and gastrointestinal activities and may be useful  
 CC in preparing a composition for diagnosing or treating a disease.  
 CC particularly a setrone-treatable disease. Such a disease or dysregulation  
 CC is related to the central and peripheral nervous system or secondary to  
 CC drug treatment, such as postoperative nausea and/or vomiting, chronic  
 CC fatigue syndrome, obsessive-compulsive disorders, schizophrenia,  
 CC alcoholism, anxiety disorders, tourette syndrome, migraine, headache and  
 CC gastrointestinal motility disorders, preferably nausea and/or vomiting  
 CC secondary to cancer chemotherapy. The polynucleotides and polypeptides  
 CC may also be useful as forensic markers. The current sequence is that of  
 CC the human HTR3B gDNA of the invention.

XX Sequence 124167 BP; 32696 A; 28609 C; 27970 G; 34892 T; 0 U; 0 Other;

Query Match 10.1%; Score 202.6; DB 12; Length 124167;

Best Local Similarity 84.3%; Pred. No. 2.2e-40; Matches 264; Conservative 0; Mismatches 44; Indels 5; Gaps 3;

QY 1216 TCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTGTACCCAGGCTGAGTACA 1275  
 DB 16571 TTTTTTTTTTTTTTTTTTTTGTAGAGGAGTCTCACTTGTAGCCAGGCTGAGTACA 16630

QY 1276 GTGGCTGTGTCTCGGCTACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG-- 1333  
 DB 16631 GTGG-TGGCATCTCAGCTACCAACAACCTCTGCTCTGTGTTCAAGGATTTCTCTGCC 16689

QY 1334 TCAGCTTCCGAATAGCTGGGATTAAGGCGCATG--ACACCATGCTTAATTTTGTGA 1391  
 DB 16690 TCAGCTTCTTGAGTAACTGGGATTAAGGATGCTGCGCACCATACCTGGCTTAATTTTGTGA 16749

QY 1392 TTTTATAGTAGACAGAGTTTGGCATGTTGACAGGCTTGCCTTGAATCTCTGACTTCA 1451



```
Db      16750  TTTTATGAGACCGGGTTTTCACCATGTGGCCAGGCTGGGCTTGAACCTCCGACTCA 16809
Oy      1452  GGTATTCACCCACCTGACCCCTCCAAAGACCTGGATTACAGGATGAGCAGCCGTGCC 1511
Db      16810  GGTATTCACCCACCTGCTCTCCAAAGTGTGGATTACAGGATGAGCAGCTGCCCC 16869
Oy      1512  CAGCTGTTTCT 1524
Db      16870  CAGCTGTACT 16882

RESULT 29
AAK70510
ID      AAK70510 standard; DNA; 6670 BP.
XX
XX      AAK70510;
AC
XX      06-NOV-2001 (first entry)
DT
XX
DE      Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25322.
XX
KM      Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM      cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS      Homo sapiens.
XX
PM      MO200157182-A2.
PD
XX      09-AUG-2001.
XX
XX      17-JAN-2001; 2001WO-US001354.
PF
XX
XX      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205515P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      11-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225477P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226799P.
PR      22-AUG-2000; 2000US-0226811P.
PR      22-AUG-2000; 2000US-0226868P.
PR      22-AUG-2000; 2000US-0227182P.
PR      23-AUG-2000; 2000US-0227009P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.

PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231242P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      14-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241221P.
PR      20-OCT-2000; 2000US-0241785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
PR      01-NOV-2000; 2000US-0244617P.
PR      08-NOV-2000; 2000US-0246474P.
PR      08-NOV-2000; 2000US-0246475P.
PR      08-NOV-2000; 2000US-0246476P.
PR      08-NOV-2000; 2000US-0246477P.
PR      08-NOV-2000; 2000US-0246478P.
PR      08-NOV-2000; 2000US-0246523P.
PR      08-NOV-2000; 2000US-0246524P.
PR      08-NOV-2000; 2000US-0246525P.
PR      08-NOV-2000; 2000US-0246526P.
PR      08-NOV-2000; 2000US-0246527P.
PR      08-NOV-2000; 2000US-0246528P.
PR      08-NOV-2000; 2000US-0246532P.
PR      08-NOV-2000; 2000US-0246609P.
PR      08-NOV-2000; 2000US-0246610P.
PR      08-NOV-2000; 2000US-0246611P.
PR      08-NOV-2000; 2000US-0246613P.
PR      17-NOV-2000; 2000US-0249207P.
PR      17-NOV-2000; 2000US-0249208P.
PR      17-NOV-2000; 2000US-0249209P.
PR      17-NOV-2000; 2000US-0249210P.
PR      17-NOV-2000; 2000US-0249211P.
PR      17-NOV-2000; 2000US-0249212P.
PR      17-NOV-2000; 2000US-0249213P.
PR      17-NOV-2000; 2000US-0249214P.
PR      17-NOV-2000; 2000US-0249215P.
```

PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclosure; SEQ ID NO 25322; 3071bp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen; (1)  
CC amino acid sequences given in AAM82170 to AAM91921. (1) have cytosolic  
CC activity, and can be used in gene therapy and vaccine production. (1)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (1) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (1) by expressing inactive proteins or to  
CC supplement the patient's own production of (1). Additionally, (1)  
CC polynucleotides may be used to produce the secreted (1), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (1) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX Sequence 6670 BP; 1577 A; 1614 C; 1756 G; 1720 T; 0 U; 3 Other;  
SQ

Query Match 10.1%; Score 202.4; DB 4; Length 6670;  
Best Local Similarity 85.3%; Pred. No. 7.4e-41;  
Matches 261; Conservative 0; Mismatches 41; Indels 4; Gaps 3;

QY 1216 TCTTTTATTTATTTTATTTTGAACAGAGTCTGACCTTGACCCAGCGGTGAGTAA 1275  
DB 2519 TTTTATTTTATTTTATTTTGAAGCAAGTTTACCTGTGACCCAGCGGTGAGTAA 2578  
QY 1276 GTGGCTGTGTCGTGGCTCAGTCAACCTCTGCTCCAGATTCAAGCATTTCTCTG-- 1333  
DB 2579 GTGGC-ACGATCTTGAGCTCAGTCAACCTCTGCTCCAGATTCAAGCATTTCTCTG 2637  
QY 1334 TTAGCTTCCGGAATAGTGTGATTAAGGCGCATGACAC-CATGCCATATTTTGTAT 1392  
DB 2638 TTAGCTTCCGGAATAGTGTGATTAAGGCGCATGACAC-CATGCCATATTTTGTAT 2697  
QY 1393 TTTTAGTAGACAGAGATTCCCATGTTGACCAAGGCTTGCAACTCTGACTTCAAG 1452  
|||||

DB 2698 TTTTAGTAGACAGAGGCGGTTTCACATGTGGCCAGGCTTTTCTCGAACTCTGACCTTAG 2757  
QY 1453 GTGATCCACCCAGCTCAGCTCCCAAGCATCGGAGTTACAGGCACTGAGCCACCGTCC 1512  
DB 2758 GTGATCCGCGCCAGCTCAGCTCCCAAGCATCGGAGTTACAGGCACTGAGCCACCA 2817  
QY 1513 AGCCTG 1518  
DB 2818 GGCCTG 2823  
RESULT 30  
AEA61181/C  
ID AEA61181 standard; DNA; 64829 BP.  
XX  
XX AEA61181;  
AC  
XX  
XX 25-AUG-2005 (first entry)  
DT  
XX  
XX Human FHL1 gene genomic sequence SEQ ID NO:91.  
DE  
XX  
XX DNA methylation; biomarker; cancer; gene; ds; FHL1.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX US2005130172-A1.  
PN  
XX  
XX 16-JUN-2005.  
PD  
XX  
XX 27-JAN-2004; 2004US-00765790.  
PF  
XX  
XX 16-DEC-2003; 2003US-00737082.  
PR  
XX  
XX (FARB ) BAYER CORP.  
PA  
XX  
XX Beard C, Burgess C, Cannon A, Harvey J, Lechner JF, Li Z;  
PI WPI. 2005-456891/46.  
XX  
XX GENBANK; AF098518, AF220153, U29538, NM\_001449.  
DR  
XX  
XX Identifying nucleic acid sequences as biomarker for disease, by  
PT identifying nucleic acid sequences comprising methylated CpG site and  
PT down-regulated in diseased cells and comparing its expression level with  
PT demethylated nucleic acid.  
XX  
XX Claim 11; SEQ ID NO 91; 27bp; English.  
PS  
XX  
XX The invention relates to a method (M1) for identifying one or more  
CC nucleic acid sequences useful as a biomarker for a disease to be  
CC detected. (M1) involves identifying nucleic acid sequences comprising  
CC methylated CpG site in promoter-first exon region and that are down-  
CC regulated in diseased cells, comparing expression level of nucleic acid  
CC sequences with that of demethylated nucleic acid sequences and  
CC identifying nucleic acid sequences exhibiting increase in expression  
CC after demethylation. Also described: (1) detecting (M2) the presence or  
CC stage of a disease in a subject, which involves determining the degree of  
CC methylation of one or more CpG sites on nucleic acid sequences in a  
CC biological sample obtained from the subject, and determining the presence  
CC of, predisposition to, or stage of the disease in the subject based on  
CC the degree of methylation; (2) monitoring the onset, progression, or  
CC regression of a disease in a subject; (3) determining the efficacy of a  
CC test compound for inhibiting a disease in a subject; and (4) a kit (1)  
CC useful for diagnosis, prognosis, staging, monitoring, and therapeutic  
CC treatment of a disease. (M1) is useful for identifying one or more  
CC nucleic acid sequences useful as a biomarker for a disease to be  
CC detected, where the nucleic acid sequences are useful for detecting, the  
CC presence or stage of a disease such as cancer e.g. colorectal cancer in a  
CC subject. The present sequence represents a specifically claimed human  
CC genomic sequence for use in the method of the invention. Note - The  
CC sequence data for this patent is not represented in the printed  
CC specification but was obtained in electronic format from the USPRO web  
XX site.

SQ Sequence 64829 BF, 17935 A, 13068 C, 14206 G, 19620 T, 0 U, 0 Other;  
Query Match 10.1%; Score 202.4; DB 14; Length 64829;  
Best Local Similarity 77.7%; Pred. No. 1.9e-40;  
Matches 283; Conservative 0; Mismatches 76; Indels 5; Gaps 3;  
QY 1192 AGTATTGATTAAGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCTCA 1251  
DB 36622 ATTTCTGTTTTTTTTTTGTTGTTGTTTTTTTGTGTTTATTTTGAAGACAGTCTCG 36563  
QY 1252 CTTTGTCCACCGAGGTGAGTACAGTACGTGCTGCGCTCAGCTGCAACTGCGCTC 1311  
DB 36562 CTCTGTCCGCCAGGCTGAGTGCAGTGGC-GCAATCTTGCTCAGCTGCAACTGCGCTC 36504  
QY 1312 CCAGGTTGACGAGTTCCTCTG--TCAGCTCCGAAATAGCTGGATTACAGGCCATGC 1369  
DB 36503 CCGAGTTCAAGGATTCCTCTCAGCTCCCAAGTACGCTGGATTACAGGCCATGC 36444  
QY 1370 ACCACCATGC--CTAATTTTGTATTTTATTTAGAGACAGAGTTTCCGATGACCG 1427  
DB 36443 CATCAGTTCAAGCTAATTTTGTATTTTATTTAGTAGAGACAGGGTTTCGATGTCATG 36384  
QY 1428 GCTTCCCTTGAACCTCTACTAGCTTACAGTATTCACCCACTGCTCCCAAGCACTGG 1487  
DB 36383 GCTGCTCTGAACTTCTGACCTCAGTATTCACCCACTGCTCCCAAGTGCAGGT 36324  
QY 1488 ATTACAGCATGAGCAGCCGTCGCCAGCTGTTTCTCAGATCTGATTTGTTCTGAA 1547  
DB 36323 ATTACAGCGGTATGACCATCGCCGCCAGCTCTTGTCTTATACATTAATTTTCATGA 36264  
QY 1548 GCCT 1551  
DB 36263 CCGT 36260  
RESULT 31  
AAS36758  
ID AAS36758 standard; DNA, 12149 BP.  
XX  
AC AAS36758;  
DT 17-DEC-2001 (first entry)  
XX  
DE Human cardiovascular system antigen genomic DNA SEQ ID No 2258.  
XX  
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;  
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;  
KW antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;  
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; tissue regeneration;  
KW anti-fertility.  
XX  
OS Homo sapiens.  
XX  
XX WO200155321-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001340.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225279P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227109P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0228287P.  
PR 01-SEP-2000; 2000US-0228343P.  
PR 01-SEP-2000; 2000US-0228344P.  
PR 01-SEP-2000; 2000US-0228345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0239340P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.

Query	DB	Score	Length	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	DB 380	DB 381	DB 382	DB 383	DB 384	DB 385	DB 386	DB 387	DB 388	DB 389	DB 390	DB 391	DB 392	DB 393	DB 394	DB 395	DB 396	DB 397	DB 398	DB 399	DB 400	DB 401	DB 402	DB 403	DB 404	DB 405	DB 406	DB 407	DB 408	DB 409	DB 410	DB 411	DB 412	DB 413	DB 414	DB 415	DB 416	DB 417	DB 418	DB 419	
-------	----	-------	--------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--

XX US2003059908-A1.  
PN  
XX  
PD 27-MAR-2003.  
XX  
PF 07-MAR-2002; 2002US-00091504.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205151P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226811P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229309P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-023297P.  
PR 14-SEP-2000; 2000US-023298P.  
PR 14-SEP-2000; 2000US-023299P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-023497P.  
PR 25-SEP-2000; 2000US-023498P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.

PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251865P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-00759678P.  
PR 17-JAN-2001; 2001US-00764869.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX	Rosen CA, Ruben SM, Barash SC;
XX	WPI; 2003-743766/70.
XX	New cardiovascular system related polynucleotides and polypeptides,
XX	PT useful for preventing, treating, or ameliorating a medical condition,
XX	PT such as cancer of cardiovascular tissues and cancer metastases.
XX	Claim 1; SEQ ID NO 2258; 262bp; English.
XX	The invention relates to human cardiovascular system related polypeptides
XX	CC and the polynucleotides encoding them. The polypeptides, polynucleotides
XX	CC and antibodies to the polypeptides are useful for diagnosing a
XX	CC pathological condition or a susceptibility to a pathological condition,
XX	CC for preventing, treating, or ameliorating a medical condition, such as
XX	CC cancer of cardiovascular system tissues, proliferative disorders, foetal
XX	CC and developmental abnormalities, haematopoietic disorders, diseases of
XX	CC the immune system, AIDS, autoimmune diseases (e.g., rheumatoid
XX	CC arthritis), inflammation, allergies, neurological disorders (e.g.,
XX	CC Alzheimer's disease, Parkinson's disease), cognitive disorders,
XX	CC schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis,
XX	CC diabetes, atherosclerosis, cardiovascular disorders, angiogenic
XX	CC disorders, kidney disorders, gastrointestinal disorders, pregnancy-
XX	CC related disorders, endocrine disorders and infections. The nucleic acids
XX	CC are also useful for chromosome identification, radiation hybrid mapping
XX	CC or long-range restriction mapping. The polypeptides and polynucleotides
XX	CC may also be used as food additives or preservatives to increase or
XX	CC decrease storage capabilities, fat content or other nutritional
XX	CC components. This sequence represents human cardiovascular system related
XX	CC genomic DNA of the invention.
XX	Sequence 12149 BP; 2788 A; 3327 C; 3233 G; 2801 T; 0 U; 0 Other;
XX	Query Match 10.1%; Score 202.2; DB 10; Length 12149;
XX	Best Local Similarity 76.4%; Pred. No. 1.1e-40;
XX	Matches 288; Conservative 0; Mismatches 83; Indels 6; Gaps 3;
QY	1196 TTGTGATATGTTTTCACATCTTTTATTTATTTTGTGAAACAGTCTACCTT 1255
DB	9764 TTTTGTGTTTGTGTTGTTGTTTGTGTTTGTGTTTGTGAGACAGTCTACCTT 9823
QY	1256 GTCAACCAAGCTGAGTACAGTGTGTGCTCGGCTACGCAACCTCTGCTCCAG 1315
DB	9824 GTCAACCAAGCTGAGTACAGTGTGTGCTCGGCTACGCAACCTCTGCTCCAG 9882
QY	1316 GTTCAAGCATTTCTCTG--TCAGCTTCCGATATAGCTGGATTAACAGCGCATG---CA 1370
DB	9883 GTTCAACGATTTCTCTGCTCAGCTCTGAGTACCTGGATTAACAGCGCATG---CA 9942
QY	1371 CCACCATGCTTAATTTTGTATTTTGTAGAGACAGAGTTGGCATTTGAACAGGCT 1430
DB	9943 AGCGCTGGTAAATTTTGTATTTTGTAGAGAGATGGGCTTTGCGCATTTGGCCAGGCT 10002
QY	1431 TGCCTTGAATCTCTAGCTCAGGTGATTCACCAACCTCAGCCCTCCAAAGCACTGGGATT 1490
DB	10003 GGTCTCAACTCTCTGCTCAGGTGATTCAGCTCAGCCCTCCAAAGTCTGGGATT 10062
QY	1491 ACAGGATAGGACAGCGTCCAGCGCTGTTTCTCAGATCTGTAATTTGTTCTGAAGCC 1550
DB	10063 ACAGGATAGGACAGCGCTGTCAGAGCTTTCTTCTTCTTCTTCCATTTTCTT 10122
QY	1551 TTCAATTTCTATCTTCTT 1567
DB	10123 TCCTTTCTTCTTCTT 10139

RESULT 33  
ADJ08870  
ID ADJ08870 strand: DNA; 12149 BP.  
AC ADJ08870;  
XX

DT	04-NOV-2004 (first entry)
XX	Human cardiovascular system associated polypeptide-related DNA SeqID2258.
DE	autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
XX	breast neoplasms; liver neoplasm; cardiovascular disorder;
XX	KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia;
XX	KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
XX	KW ocular disorder; corneal infection; wound healing;
XX	KW epithelial cell proliferation; skin aging; sunburn;
XX	KW organ transplantation; cell culture; tissue regeneration; chemotaxis;
XX	KW food additive; preservative; cardiovascular system associated antigen;
XX	KW nuclear factor kappaB; NFkappaB; promoter element; human; ds.
OS	Homo sapiens.
XX	US2004005575-A1.
PN	08-JAN-2004.
XX	26-AUG-2002; 2002US-00227577.
XX	31-JAN-2000; 2000US-0179065P.
XX	04-FEB-2000; 2000US-0180628P.
XX	24-FEB-2000; 2000US-0184664P.
XX	02-MAR-2000; 2000US-0186350P.
XX	16-MAR-2000; 2000US-0189874P.
XX	17-MAR-2000; 2000US-0190076P.
XX	18-APR-2000; 2000US-0198123P.
XX	19-MAY-2000; 2000US-0205515P.
XX	07-JUN-2000; 2000US-0209467P.
XX	28-JUN-2000; 2000US-0214886P.
XX	30-JUN-2000; 2000US-0215135P.
XX	07-JUL-2000; 2000US-0216647P.
XX	07-JUL-2000; 2000US-0216880P.
XX	11-JUL-2000; 2000US-0217487P.
XX	11-JUL-2000; 2000US-0217496P.
XX	14-JUL-2000; 2000US-0218290P.
XX	26-JUL-2000; 2000US-0220963P.
XX	26-JUL-2000; 2000US-0220964P.
XX	14-AUG-2000; 2000US-0224518P.
XX	14-AUG-2000; 2000US-0224519P.
XX	14-AUG-2000; 2000US-0225213P.
XX	14-AUG-2000; 2000US-0225214P.
XX	14-AUG-2000; 2000US-0225266P.
XX	14-AUG-2000; 2000US-0225267P.
XX	14-AUG-2000; 2000US-0225268P.
XX	14-AUG-2000; 2000US-0225270P.
XX	14-AUG-2000; 2000US-0225447P.
XX	14-AUG-2000; 2000US-0225757P.
XX	14-AUG-2000; 2000US-0225758P.
XX	14-AUG-2000; 2000US-0225759P.
XX	18-AUG-2000; 2000US-0226279P.
XX	22-AUG-2000; 2000US-0226811P.
XX	22-AUG-2000; 2000US-0226868P.
XX	22-AUG-2000; 2000US-0227182P.
XX	23-AUG-2000; 2000US-0227039P.
XX	30-AUG-2000; 2000US-0228924P.
XX	01-SEP-2000; 2000US-0229287P.
XX	01-SEP-2000; 2000US-0229343P.
XX	01-SEP-2000; 2000US-0229344P.
XX	01-SEP-2000; 2000US-0229345P.
XX	05-SEP-2000; 2000US-0229509P.
XX	05-SEP-2000; 2000US-0229513P.
XX	06-SEP-2000; 2000US-0230437P.
XX	06-SEP-2000; 2000US-0230438P.
XX	08-SEP-2000; 2000US-0231242P.
XX	08-SEP-2000; 2000US-0231243P.
XX	08-SEP-2000; 2000US-0231244P.
XX	08-SEP-2000; 2000US-0231413P.
XX	08-SEP-2000; 2000US-0231414P.
XX	08-SEP-2000; 2000US-0232080P.
XX	08-SEP-2000; 2000US-0232081P.





DB 10003 GGTCTCAACTCTGCTGCTCAGGTGATCCGCCACTCAGCCCTCCAAAGTGTGGGATT 10062  
QY 1491 ACAGGATAGACCGGTCGCCAGCTGTTTCTCAGATCCGTATTTGTTCTGAAGC 1550  
DB 10063 ACAGGATAGACCGGTCGCTGCTCAGTCTTTCTTTCTTTCCATTTTCCCTT 10122  
QY 1551 TTCATTTCTATCTTCTT 1567  
DB 10123 TCCTTCTTCTTCTTCTT 10139

RESULT 34  
AEB32377/c  
ID AEB32377 standard; DNA; 160274 BP.  
AC AEB32377;  
XX  
XX 08-SEP-2005 (first entry)  
DE Human genomic DNA #18.  
XX  
XX SNP detection; diagnosis; non-insulin dependent diabetes; obesity;  
KW antidiabetic; anorectic; endocrine disease; gastrointestinal disease;  
KM metabolic disorder; nutritional disorder; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX US2005147987-A1.  
PN  
XX 07-JUL-2005.  
PD  
XX 19-JUL-2004; 2004US-00893315.  
PF  
XX 08-SEP-2000; 2000US-0231397P.  
PR 10-SEP-2001; 2001US-00948947.  
XX  
XX (APPL-) APPLERA CORP NY.  
PA  
PI Venter JC, Zhang JN, Liu X, Rowe W, Cravchik A, Kalush F;  
PI Naik A, Subramanian G, Woodage T;  
PI WPI; 2005-511776/52.  
DR  
XX  
XX New detection reagent capable of detecting 1, 100, 500, 1000 or 5000 or  
PT more single nucleic acid polymorphisms, useful in identifying an  
PT individual having or at risk of developing type II diabetes or obesity.  
XX  
XX Disclosure; SEQ ID NO 140; 31pp; English.

CC The invention relates to a detection reagent capable of detecting one or  
CC more single nucleic acid polymorphisms. The invention also relates to  
CC determining whether a trait is linked to one of the human chromosomes or  
CC its sub-region, a computer readable medium having stored in it the SNP  
CC relational information given in the specification, an isolated nucleic  
CC acid molecule for detecting at least one SNP given in the specification  
CC comprising at least about 12 contiguous nucleotides, genotyping at least  
CC one SNP position given in the specification in a sample, identifying an  
CC individual having or at risk of developing a disorder and a kit  
CC comprising at least one container containing a detection reagent.  
CC Determining whether a trait is linked to one of the human chromosomes or  
CC its sub-region comprises determining whether the trait is linked to one  
CC or more SNPs using the detection reagents. Genotyping at least one SNP  
CC position given in the specification in a sample comprises contacting the  
CC sample with a detection reagent that differentiates between alternative  
CC alleles at at least one SNP position given in the specification, and  
CC determining which allele is present at the at least one SNP position.  
CC Identifying an individual having or at risk of developing a disorder  
CC comprises genotyping at least one SNP given in the specification in a  
CC nucleic acid sample from the individual. The disorder is type II diabetes  
CC (non-insulin dependent diabetes) or obesity. The detection reagent is  
CC useful in identifying an individual having or at risk of developing a  
CC disorder, particularly type II diabetes or obesity. This sequence  
CC represents human genomic DNA used in the scope of the invention. Note:

CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 160274 BP; 50545 A; 28535 C; 28704 G; 52490 T; 0 U; 0 Other;  
Query Match 10.1%; Score 202; DB 14; Length 160274;  
Best Local Similarity 77.2%; Pred. No. 3 5e-40;  
Matches 284; Conservative 0; Mismatches 80; Indels 4; Gaps 3;

QY 1199 GATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGAGTCTGACTTGTG 1258  
DB 21584 GATAGCTTATTTATTTATTTATTTATTTATTTTATTTTGAACAGAGTCTGACTTGTG 21525  
QY 1259 ACCGAGCTGAGTACAGTGTGCTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1318  
DB 21524 ACCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21466  
QY 1319 CAGGCAATTCCTCTG--TCAGCTTCCGAAATAGCTGGAATTCAGAGCGCATGCAACAC-C 1375  
DB 21465 CAGCAATTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21406  
QY 1376 ATGCTTAAATTTTGTATTTTATTTAGTACAGAGTGTGCTGCTGCTGCTGCTGCTGCT 1435  
DB 21405 TAGGCTAAATTTTGTATTTTATTTAGTACAGAGTGTGCTGCTGCTGCTGCTGCTGCT 21346  
QY 1436 TGAATCTCTGATTTGATGATGATCCACCACTGACCTCCCAAGCACTGGGATTTACAG 1495  
DB 21345 TGAATCTCTGATTTGATGATGATCCACCACTGACCTCCCAAGCACTGGGATTTACAG 21286  
QY 1496 CATGAGCCACCGTGGCCAGCGCTGTTTCTCAGATCTGATTTGTTTGAAGCTTCAT 1555  
DB 21285 TATGATCCACCAAGCTGATGCTTATTTTATTTTATTTTATTTTCAACAGTTAAT 21226  
QY 1556 TTCTATCT 1563  
DB 21225 TAGAGTCT 21218

RESULT 35  
AEB32388/c  
ID AEB32388 standard; DNA; 160300 BP.  
XX  
XX AEB32388;  
XX  
XX 08-SEP-2005 (first entry)  
DE Human genomic DNA #29.  
XX  
XX SNP detection; diagnosis; non-insulin dependent diabetes; obesity;  
KW antidiabetic; anorectic; endocrine disease; gastrointestinal disease;  
KM metabolic disorder; nutritional disorder; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX US2005147987-A1.  
PN  
XX 07-JUL-2005.  
PD  
XX 19-JUL-2004; 2004US-00893315.  
PF  
XX 08-SEP-2000; 2000US-0231397P.  
PR 10-SEP-2001; 2001US-00948947.  
XX  
XX (APPL-) APPLERA CORP NY.  
PA  
PI Venter JC, Zhang JN, Liu X, Rowe W, Cravchik A, Kalush F;  
PI Naik A, Subramanian G, Woodage T;  
PI WPI; 2005-511776/52.  
DR  
XX  
XX New detection reagent capable of detecting 1, 100, 500, 1000 or 5000 or  
PT more single nucleic acid polymorphisms, useful in identifying an





```
Db      11505 TCACACCCACCTCGCGCTCCCAAGTGTGGGATTACAGGCATGACCAACCGCGCCGACC 11446
Oy      1517 TGTTCCTCAGATCCTGTATTGTTCT 1544
          | | | | | | | | | |
Db      11445 TAACTTACTCTTCTTGAAGCTGCTTTT 11418

RESULT 37
ADK66161/c
ID      ADK66161 standard; DNA; 51256 BP.
Ac      ADK66161;
Df      06-MAY-2004 (first entry)
Df      Human protease gene.
XX      Protease; therapeutic; drug; diagnosis; pharmacogenomic analysis;
KW      treatment; human; chromosome 3; gene; ds.
XX      Homo sapiens.
OS
XX
Ph      Key
FT      Location/Qualifiers
FT      replace(425,G)
FT      /tag= a
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(1521..1522,TG)
FT      /tag= b
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(2245,A)
FT      /tag= c
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(2492,G)
FT      /tag= d
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      /tag= e
FT      /product= "Human protease protein"
FT      3000..4512
FT      /tag= f
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      CDS
FT      /tag= g
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      exon
FT      /tag= h
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      4513..17481
FT      /tag= i
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= j
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= k
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= l
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= m
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= n
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= o
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= p
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= q
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= r
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= s
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= t
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= u
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= v
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= w
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= x
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= y
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= z
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= aa
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ab
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ac
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ad
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ae
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= af
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ag
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ah
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ai
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= aj
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ak
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= al
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= am
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= an
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ao
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      variation
FT      /tag= ap
```

```
FT intron 17588..17739
FT /tag= ag
FT replace(17634,A)
FT variation
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 17740..17878
FT /tag= as
FT intron 17879..18180
FT /tag= at
FT replace(18033,G)
FT /tag= au
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 18181..18228
FT /tag= av
FT intron 18229..21348
FT /tag= aw
FT replace(18280,C)
FT /tag= ax
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18301,T)
FT /tag= ay
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18306,A)
FT /tag= az
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18458,A)
FT /tag= ba
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(20037,G)
FT /tag= bb
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 21349..21426
FT /tag= bc
FT intron 21427..41548
FT /tag= bd
FT replace(21591,A)
FT /tag= be
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(22083,A)
FT /tag= bf
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(22258,G)
FT /tag= bg
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(23228,T)
FT /tag= bh
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(25434..25436,AA)
FT /tag= bi
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(25435,A)
FT /tag= bj
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(25869,G)
FT /tag= bk
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation
FT /tag= bl
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT /tag= bm
FT variation
FT replace(26048,C)
FT /tag= bn
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(26191,A)
FT variation

Query Match 10.1%; Score 201.6; DB 10; Length 51256;
Best Local Similarity 82.0%; Pred. No. 2.8e-40;
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;

QY 1221 TTAATTTTATTTTGAACAAGAGTCTCACTTTGTACCCAGGCTGAGTACAGTGGC 1280
DB 11744 TTTTATTTTATTTTGAAGAGTCTCACTTTGTACCCAGGCTGAGTACAGTGGC 11686
QY 1281 TGTGCTCTGGCTCACTGCAACCTCTGCTCCAGGTTGAAGGATTTCCTG--TCAC 1338
```

```
DB 11685 TCGATCTCAACTCACTGCAACCTCCAACTCAAGTACAAAGTAATTCCTGCTCAGC 11626
QY 1339 TTCCGAAATAGCTGGATTTACAGGGGATGC--ACCAATCATGCTTAATTTTGTATTTT 1396
DB 11625 CTCCCAAGTAGCTGGGACTACAGGGGCAATGCCACATGCTGCTTAATTTTGTATTTT 11566
QY 1397 AGTAGAGACAGATTTCGCGATGTTGACCAAGGCTTGCTGAATCTGACTTCAGGTGA 1456
DB 11565 AGTAGAGACAGATTTCACCAATTTGTCAGGCTTAATCTTGAATCTTCACTTCAGGTGA 11506
QY 1457 TCACCCCACTGAGCTCCCAAGCACTGGATTACAGGCATGAGCCACCGTCCAGCC 1516
DB 11505 TCACCCCACTGAGCTCCCAAGCACTGGATTACAGGCATGAGCCACCGTCCAGCC 11446
QY 1517 TGTTTTCTCAGATCTCTGATTTGTTCT 1544
DB 11445 TAACTTACTCTCTTGAAGCTGCTTTT 11418
```

```
RESULT 38
ABD3611/C
ID ABD3611 standard; DNA; 89567 BP.
XX
AC ABD3611;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-124.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
XX of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 838; 1822p; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 89567 BP; 24453 A; 17570 C; 19182 G; 28362 T; 0 U; 0 Other;
```

Query Match	10.1 %;	Score 201.6;	DB 13;	Length 89567;
Best Local Similarity	84.3 %;	Pred. No. 3.5e-40;		
Matches 263;	Conservative 0;	Mismatches 44;	Indels 5;	Gaps 3;
QY 1217	CTTTTATTTTATTTTATTTTATTTTGAACAGAGTCTCAGCTTGTCACCCAGGCTGGAGTACAG			1276
Db 37961	CTGTTTTTTGTTTTTTTTTTTTTTGAGACAGAGTCTCAGTGTCTCTCCAGGCTGGAGTACAG			37902
QY 1277	TGGCTGTGCTCGGCTCAGTCGAACCTGTGCTCCAGGTTCAAGCAATTCCTG--T			1334
Db 37901	TGG-TGCAATCTGGATTCACTGCAACCTGTGCTCTCGGAGTTCAAGCAATTCCTGCTCT			37843
QY 1335	CAGCTTCCGAATAGCTGGAGATTACAGGGCCATGC--ACACATAGCCAAATTTTTGAT			1392
Db 37842	CAGCTCTCTAGTAGAGCTGGAGATTACAGTGCATAGCCATCATGCCCAGCTAAATTTTTGAT			37783
QY 1393	TTTTAGTAGAGACAGAGTTTCGGCATTTTGACCAAGGCTTGCCTTGAATCTGTGACTTTCAG			1452
Db 37782	TTTTAGTAGAGACAGGGGTTTTCAACCATTTTGCCAGGCTGTGTTCAAACTCTGACCCGAG			37723
QY 1453	GTGATCCACCCACCTCAGCTCCCAAGAAGCACTGGGATTACAGGCATGAGCAACCGTGCC			1512
Db 37722	ATGATCCACCCACCTCAGCTCCCAAGAAGGTGTGGGATTACAGGCGTGAAGCACCATGGCC			37663
QY 1513	AGCCTGTTTTCT	1524		
Db 37662	AGCCTCCTGTCT	37651		

RESULT 39	
ADe43582	
ID	ADe43582 standard; DNA; 128034 BP.
XX	
AC	ADe43582;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Polymorphic human IDE genomic sequence, SEQ ID 187.
XX	
KW	Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;
KM	Alzheimer's disease; neuroprotective; neurotropic; gene therapy;
KX	Chromosome 10; gene; ds.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	misc_feature
FT	2456
FT	/*tag= a
FT	/note= "There is a variation at this position"
FT	3279
FT	/*tag= b
FT	/note= "There is a variation at this position"
FT	3407
FT	/*tag= C
FT	/note= "There is a variation at this position"
FT	17095
FT	/*tag= d
FT	/note= "There is a variation at this position"
FT	17242
FT	/*tag= e
FT	/note= "There is a variation at this position"
FT	33590
FT	/*tag= f
FT	/note= "There is a variation at this position"
FT	38903
FT	/*tag= g
FT	/note= "There is a variation at this position"
FT	42943
FT	/*tag= h
FT	/note= "There is a variation at this position"
FT	43391
FT	/*tag= i
FT	/note= "There is a variation at this position"

PT	misc_feature	45017	/*tag= j	"There is a variation at this position"
FT		/note=		
FT	misc_feature	62458	/*tag= k	"There is a variation at this position"
FT		/note=		
FT	misc_feature	68906	/*tag= l	"There is a variation at this position"
FT		/note=		
FT	misc_feature	68973	/*tag= m	"There is a variation at this position"
FT		/note=		
FT	misc_feature	69586	/*tag= n	"There is a variation at this position"
FT		/note=		
FT	misc_feature	73772	/*tag= o	"There is a variation at this position"
FT		/note=		
FT	misc_feature	74084	/*tag= p	"There is a variation at this position"
FT		/note=		
FT	misc_feature	83024	/*tag= q	"There is a variation at this position"
FT		/note=		
FT	misc_feature	83104	/*tag= r	"There is a variation at this position"
FT		/note=		
FT	misc_feature	89301	/*tag= s	"There is a variation at this position"
FT		/note=		
FT	misc_feature	10560	/*tag= t	"There is a variation at this position"
FT		/note=		
FT	misc_feature	107395	/*tag= u	"There is a variation at this position"
FT		/note=		
FT	misc_feature	108489	/*tag= v	"There is a variation at this position"
FT		/note=		
FT	misc_feature	111914	/*tag= w	"There is a variation at this position"
FT		/note=		
FT	misc_feature	112114	/*tag= x	"There is a variation at this position"
FT		/note=		
FT	misc_feature	113142	/*tag= y	"There is a variation at this position"
FT		/note=		
FT	misc_feature	11351	/*tag= z	"There is a variation at this position"
FT		/note=		
FT	misc_feature	11463	/*tag= aa	"There is a variation at this position"
FT		/note=		
FT	misc_feature	11662	/*tag= ab	"There is a variation at this position"
FT		/note=		
FT	misc_feature	117803	/*tag= ac	"There is a variation at this position"
FT		/note=		
FT	misc_feature	124565	/*tag= ad	"There is a variation at this position"
FT		/note=		
XX	WO2003054143-A2.			
XX				
PD	03-JUL-2003.			
XX				
PF	25-OCT-2002; 2002WO-US034679.			
XX				
PR	25-OCT-2001; 2001US-0339525P.			
PR	08-NOV-2001; 2001US-0336929P.			
PR	08-NOV-2001; 2001US-0338010P.			

PR 09-NOV-2001; 2001US-0338363P.  
PR 04-DEC-2001; 2001US-0337052P.  
PR 28-MAR-2002; 2002US-0368919P.  
XX  
XX (NEUR-) NEUROGENETICS INC.  
XX (GEO) GEN HOSPITAL CORP.  
XX  
XX Becker KD, Veliceljebi G, Elliott KJ, Wang X, Tanzi RE, Bertram L;  
PI Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;  
XX WPI; 2003-559131/52.  
XX  
XX Determining a predisposition for or the occurrence of neurodegenerative  
PT disease, e.g. Alzheimer's disease by detecting in a target nucleic acid  
PT the presence or absence of an allelic variant of one or more polymorphic  
PT regions.  
XX  
XX Claim 9; Page 584-page 618; 848pp; English.  
XX  
XX The present invention relates to a method (M1) for determining a  
CC predisposition for or the occurrence of neurodegenerative disease in a  
CC subject. The method comprises detecting in a target nucleic acid obtained  
CC from the subject the presence or absence of an allelic variant of one or  
CC more polymorphic regions of one or more genes selected from uPA  
CC (urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-  
CC degrading enzyme), KNSL1 (kinesin-like protein 1), LIPA (lysosomal acid  
CC lyase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the  
CC presence of at least one of the allelic variant of one or more  
CC polymorphic regions is indicative of a predisposition for or the  
CC occurrence of neurodegenerative disease. The genes are all located on  
CC chromosome 10. M1 is useful for determining a predisposition for or the  
CC occurrence of, and for treating neurodegenerative disease, particularly  
CC Alzheimer's disease.  
XX  
XX Sequence 128034 BP; 34726 A; 25977 C; 26400 G; 40799 T; 0 U; 132 Other;  
SQ  
Query Match 10.1%; Score 201.6; DB 10; Length 128034;  
Best Local Similarity 82.0%; Pred. No. 4.1e-40;  
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;  
QY 1218 TTTTATTTTATTTTATTTTGAACAAGATCTCACTTTGTCAACCCAGGCTGAGTACACT 1277  
DB 24034 TATTCCTTTTATTTTATTTTGAAGAGATCTCACTTTGTGCCAGGCTGAGTACACT 24093  
QY 1278 GCGTGTGCTGCTGGCTCACTGCAACCTTGCTCCAGGTTGAAGGATTCTCTG--TC 1335  
DB 24094 GGC-ATAGCTTGGCTCACTGCAACCTTGCTCCAGGTTGAAGGATTCTCTGCTC 24152  
QY 1336 AGCTTCCGAATAGCTGGGATTAACAGGCGCATGCACACCA--TGCTTAATTTTGTATT 1393  
DB 24153 AGCTTCCGAATAGCTGGGATTAACAGGCGCATGCACACCACTGCTTAATTTTGTATT 24212  
QY 1394 TTTAGTAGAGACAGATTGGCATGTTGACCAAGGCTTGCCTTGAACCTCTGACTTCAG 1453  
DB 24213 TTTAATAGAGACAGATTTCACCAATGTTGGCAACCTGCTTAAGAACTCTTACTTCAG 24272  
QY 1454 TGAATCCACCACTGACCTCCCAAGCACTGGGATTACAGGCATGAGCCACGTCGCCA 1513  
DB 24272 TGAATCCACCTGACCTCCCAAGCACTGGGATTACAGGCATGAGCCACGTCGCCA 24332  
QY 1514 GCGTGTGCTGCTGGCTCACTGCAACCTTGCTCCAGGTTGAAGGATTCTCTG--TC 1541  
DB 24333 GCGTGTGCTGCTGGCTCACTGCAACCTTGCTCCAGGTTGAAGGATTCTCTG--TC 24360  
RESULT 40  
ADE43581  
ID ADE43581 standard; DNA; 128034 BP.  
XX ADE43581;  
XX AC  
XX AC  
XX DT 29-JAN-2004 (first entry)  
XX

DE Human IDE genomic sequence, SEQ ID 186.  
XX  
XX Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;  
XX Alzheimer's disease; neuroprotective; neurotrophic; gene therapy;  
XX Chromosome 10; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2003054143-A2.  
XX  
XX 03-JUL-2003.  
XX  
XX 25-OCT-2002; 2002WO-US034679.  
XX  
XX 25-OCT-2001; 2001US-0339525P.  
PR 08-NOV-2001; 2001US-0336929P.  
PR 08-NOV-2001; 2001US-0338010P.  
PR 09-NOV-2001; 2001US-0338363P.  
PR 04-DEC-2001; 2001US-0337052P.  
PR 28-MAR-2002; 2002US-0368919P.  
XX  
XX (NEUR-) NEUROGENETICS INC.  
XX (GEO) GEN HOSPITAL CORP.  
XX  
XX Becker KD, Veliceljebi G, Elliott KJ, Wang X, Tanzi RE, Bertram L;  
PI Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;  
XX WPI; 2003-559131/52.  
XX  
XX Determining a predisposition for or the occurrence of neurodegenerative  
PT disease, e.g. Alzheimer's disease by detecting in a target nucleic acid  
PT the presence or absence of an allelic variant of one or more polymorphic  
PT regions.  
XX  
XX Claim 22; Page 549-584; 848pp; English.  
XX  
XX The present invention relates to a method (M1) for determining a  
CC predisposition for or the occurrence of neurodegenerative disease in a  
CC subject. The method comprises detecting in a target nucleic acid obtained  
CC from the subject the presence or absence of an allelic variant of one or  
CC more polymorphic regions of one or more genes selected from uPA  
CC (urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-  
CC degrading enzyme), KNSL1 (kinesin-like protein 1), LIPA (lysosomal acid  
CC lyase), and TNFRSF6 (Tumour Necrosis Factor Receptor-SF6), where the  
CC presence of at least one of the allelic variant of one or more  
CC polymorphic regions is indicative of a predisposition for or the  
CC occurrence of neurodegenerative disease. The genes are all located on  
CC chromosome 10. M1 is useful for determining a predisposition for or the  
CC occurrence of, and for treating neurodegenerative disease, particularly  
CC Alzheimer's disease.  
XX  
XX Sequence 128034 BP; 34721 A; 25985 C; 26409 G; 40808 T; 0 U; 101 Other;  
SQ  
Query Match 10.1%; Score 201.6; DB 10; Length 128034;  
Best Local Similarity 82.0%; Pred. No. 4.1e-40;  
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;  
QY 1218 TTTTATTTTATTTTATTTTGAACAAGATCTCACTTTGTCAACCCAGGCTGAGTACACT 1277  
DB 24034 TATTCCTTTTATTTTATTTTGAAGAGATCTCACTTTGTGCCAGGCTGAGTACACT 24093  
QY 1278 GCGTGTGCTGCTGGCTCACTGCAACCTTGCTCCAGGTTGAAGGATTCTCTG--TC 1335  
DB 24094 GGC-ATAGCTTGGCTCACTGCAACCTTGCTCCAGGTTGAAGGATTCTCTGCTC 24152  
QY 1336 AGCTTCCGAATAGCTGGGATTAACAGGCGCATGCACACCA--TGCTTAATTTTGTATT 1393  
DB 24153 AGCTTCCGAATAGCTGGGATTAACAGGCGCATGCACACCACTGCTTAATTTTGTATT 24212  
QY 1394 TTTAGTAGAGACAGATTGGCATGTTGACCAAGGCTTGCCTTGAACCTCTGACTTCAG 1453  
DB 24213 TTTAATAGAGACAGATTTCACCAATGTTGGCAACCTGCTTAAGAACTCTTACTTCAG 24272



```
Qy 1454 TGATCCACCCAGCTCAGCCTCCCAAGCATGAGTTCAGGATGAGCCACCGTGCCCA 1513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24273 TGATCCGCTTCTCTCAGCCTCCCAAGTGTGATTCAGGTGTGAGCCACCATGCCCA 24332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1514 GCGTGTTCCTCAGATCCTGTATTGTT 1511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24333 GCGTGTTCCTTCCTTATTATT 24360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 41
ADH54059
ID ADH54059 standard; DNA, 128034 BP.
XX
XX ADH54059;
XX
XX 25-MAR-2004 (first entry)
XX
XX Human IDE gene DNA sequence SegID186.
XX
XX human; neurodegenerative disease; urokinase plasminogen activator; uPA;
XX gamma-gynuclein; SNC3; insulin degrading enzyme; IDE;
XX kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
XX tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX
XX Homo sapiens.
XX
XX OS
XX
XX Key Location/Qualifiers
XX variation replace(2456,G)
    /tag= a
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(3279,C)
    /tag= b
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(3407,T)
    /tag= c
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(17095,G)
    /tag= d
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(17242,T)
    /tag= e
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(33590,G)
    /tag= f
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(38903,T)
    /tag= g
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(42943,C)
    /tag= h
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(43391,G)
    /tag= i
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(45017,T)
    /tag= j
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(62498,C)
    /tag= k
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(68906,G)
    /tag= l
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(68973,T)
    /tag= m
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(69586,C)
    /tag= n
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(73772,G)
    /tag= o
    /standard_name= "Single nucleotide polymorphism"
XX variation replace(74084,T)
    /tag= p
    /standard_name= "Single nucleotide polymorphism"

FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(83024,T)
    /tag= q
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(83104,T)
    /tag= r
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(89301,G)
    /tag= s
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(105060,G)
    /tag= t
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(107395,A)
    /tag= u
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(108489,T)
    /tag= v
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(111914,G)
    /tag= w
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(112114,A)
    /tag= x
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(113142,G)
    /tag= y
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(113591,G)
    /tag= z
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(114683,G)
    /tag= aa
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(115662,A)
    /tag= ab
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(117803,T)
    /tag= ac
    /standard_name= "Single nucleotide polymorphism"
FT variation replace(124565,T)
    /tag= ad
    /standard_name= "Single nucleotide polymorphism"
FT variation
    /tag= ad
    /standard_name= "Single nucleotide polymorphism"
XX
XX US2003224380-A1.
XX
XX 04-DEC-2003.
XX
XX 25-OCT-2002; 2002US-00282174.
XX
XX 25-OCT-2001; 2001US-0339525P.
XX 25-OCT-2001; 2001US-0348065P.
XX 02-NOV-2001; 2001US-0335983P.
XX 08-NOV-2001; 2001US-0336929P.
XX 08-NOV-2001; 2001US-0336010P.
XX 09-NOV-2001; 2001US-0338363P.
XX 04-DEC-2001; 2001US-0337052P.
XX 28-MAR-2002; 2002US-0368919P.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Becker KD, Velicelcibi G, Elliott KJ, Wang X, Tanzi RE;
XX Berram L, Saunders AJ, Mullin KM, Sampson AJ;
XX
XX WPI; 2004-060538/06.
XX
XX Determining a predisposition for or the occurrence of neurodegenerative
XX disease, particularly Alzheimer's disease, comprises determining the
XX presence of a polymorphism in the uPA, SNC3, IDE, KNSL1, LIPA or TNFRSF6
XX gene.
XX
XX Claim 22; SEQ ID NO 186; 205pp; English.
XX
XX
```

XX This invention relates to a novel method of determining a predisposition  
CC for or the occurrence of neurodegenerative disease comprising detecting  
CC in a target nucleic acid obtained from the subject the presence of an  
CC allelic variant of polymorphic regions of human genes selected from  
CC urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin  
CC degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid  
CC lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The  
CC method is useful in determining the presence or predisposition to a  
CC neurodegenerative disease, particularly Alzheimer's disease. The present  
CC sequence is that of the human IDE gene which is related to the invention.  
SQ Sequence 128034 BP; 34731 A; 25985 C; 26409 G; 40808 T; 0 U; 101 Other;

Query Match 10.1%; Score 201.6; DB 12; Length 128034;  
Best Local Similarity 82.0%; Pred. No. 4.1e-40;  
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;

QY 1218 TTTTATTTTATTTTATTTTATTTTGAACAGAGTCTCACTTTGTACCCAGGCTGGAGTACAGT 1277  
DB 24034 TATTCCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTTGTACCCAGGCTGGAGTACAGT 24093  
QY 1278 GCGTGTGCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG--TC 1335  
DB 24094 GGC-ATAGCTTGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTGCTC 24152  
QY 1336 AGCTTCCGGAATAGCTGGATTTACAGGGCCATGACACCA--TGCTTAATTTTGTATT 1393  
DB 24153 AGCTTCCGGAATAGCTGGATTTACAGGGCCATGACACCA--TGCTTAATTTTGTATT 24212  
QY 1394 TTTAGTAGAGACAGAGTTTGGCATGTTGACCAAGGCTTGCCTTGAACCTCTGACTTCAGG 1453  
DB 24213 TTTTATAGAGACAGAGTTTACCATGTTGGCAACCTGTTCTAGAACTCTTAATCTTCAGG 24272  
QY 1454 TGATCCACCCACCTGAGCTTCCCAAGCACTG3GATTTACAGGATGAGCCACGTTGCCA 1513  
DB 24273 TGATCCGCTTCTCTGAGCTTCCCAAGTGTCTGGGATTTACAGGTTGAGCCACATGCCA 24332  
QY 1514 GCGTGTGCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG--TC 1335  
DB 24333 GCGTGTGCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG--TC 24360

## RESULT 42

ADH54060  
ID ADH54060 standard; DNA; 128034 BP.

XX AC ADH54060;  
XX DT 25-MAR-2004 (first entry)  
DE Human IDE gene variant DNA sequence SeqID187.  
XX DE  
XX human; neurodegenerative disease; urokinase plasminogen activator; uPA;  
KW gamma-synuclein; SNCG; insulin degrading enzyme; IDE;  
KW kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;  
KW tumour necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.  
OS Homo sapiens.  
XX OS  
XX US2003224380-A1.  
XX PD 04-DEC-2003.  
XX PF 25-OCT-2002; 2002US-00282174.  
XX PR 25-OCT-2001; 2001US-0339525P.  
XX PR 25-OCT-2001; 2001US-0348065P.  
XX PR 02-NOV-2001; 2001US-0336983P.  
XX PR 08-NOV-2001; 2001US-0336929P.  
XX PR 09-NOV-2001; 2001US-0338010P.  
XX PR 04-DEC-2001; 2001US-0337052P.

PR 28-MAR-2002; 2002US-0368919P.  
XX (GENO) GEN HOSPITAL CORP.  
XX PI Becker KD, Velicelbi G, Elliott KJ, Wang X, Tanzi RE;  
PI Bertram L, Saunders AJ, Mullin KM, Sampson AJ;  
DR WPI; 2004-060538/06.

XX Determining a predisposition for or the occurrence of neurodegenerative  
PT disease, particularly Alzheimer's disease, comprises determining the  
PT presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6  
PT gene.  
XX

PS Claim 9; SEQ ID NO 187; 205BP; English.

CC This invention relates to a novel method of determining a predisposition  
CC for or the occurrence of neurodegenerative disease comprising detecting  
CC in a target nucleic acid obtained from the subject the presence of an  
CC allelic variant of polymorphic regions of human genes selected from  
CC urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin  
CC degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid  
CC lipase (LIPA) and tumour necrosis factor receptor SF6 (TNFRSF6). The  
CC method is useful in determining the presence or predisposition to a  
CC neurodegenerative disease, particularly Alzheimer's disease. The present  
CC sequence is that of the human IDE gene, with polymorphic sites  
CC represented by n, which is related to the invention.  
SQ Sequence 128034 BP; 34726 A; 25977 C; 26400 G; 40799 T; 0 U; 132 Other;

Query Match 10.1%; Score 201.6; DB 12; Length 128034;  
Best Local Similarity 82.0%; Pred. No. 4.1e-40;  
Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;

QY 1218 TTTTATTTTATTTTATTTTATTTTGAACAGAGTCTCACTTTGTACCCAGGCTGGAGTACAGT 1277  
DB 24034 TATTCCTTTTATTTTATTTTATTTTGAACAGAGTCTCACTTTGTACCCAGGCTGGAGTACAGT 24093  
QY 1278 GCGTGTGCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG--TC 1335  
DB 24094 GGC-ATAGCTTGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTGCTC 24152  
QY 1336 AGCTTCCGGAATAGCTGGATTTACAGGGCCATGACACCA--TGCTTAATTTTGTATT 1393  
DB 24153 AGCTTCCGGAATAGCTGGATTTACAGGGCCATGACACCA--TGCTTAATTTTGTATT 24212  
QY 1394 TTTAGTAGAGACAGAGTTTGGCATGTTGACCAAGGCTTGCCTTGAACCTCTGACTTCAGG 1453  
DB 24213 TTTTATAGAGACAGAGTTTACCATGTTGGCAACCTGTTCTAGAACTCTTAATCTTCAGG 24272  
QY 1454 TGATCCACCCACCTGAGCTTCCCAAGCACTG3GATTTACAGGATGAGCCACGTTGCCA 1513  
DB 24273 TGATCCGCTTCTCTGAGCTTCCCAAGTGTCTGGGATTTACAGGTTGAGCCACATGCCA 24332  
QY 1514 GCGTGTGCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG--TC 1335  
DB 24333 GCGTGTGCTCGGCTCACTGCAACCTCTGCTCCAGGTTCAAGGATTTCTCTG--TC 24360

## RESULT 43

ADE4315/C  
ID ADE4315 standard; DNA; 202100 BP.

XX AC ADE4315;  
XX DT 29-JAN-2004 (first entry)  
DE Human IDE/ KNSL1 genomic sequence. SEQ ID 484.  
XX DE  
XX Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;  
KW Alzheimer's disease; neuroprotective; nootropic; gene therapy;  
KW Chromosome 10; gene; ds.  
XX



Db 99601 GGC-ATAGTCTTGCTCAGTGCACCTCTGCTCCCAAGTTCAGCAATTCCTGCTC 99543  
QY 1336 ACCTCCCGATAGCTGGATTTACAGGGCGCATGACACCA--TGCTTAATTTTGTATT 1393  
Db 99542 ACCTTCCTAGTACCGGAGCTACAGGCGTGTGCCACACCTGCTAATTTTGTATT 99483  
QY 1394 TTTAGTAGAGACAGATTTCGCGCATGTGACCAAGGCTTCTTGAATCTCTGACTTCAGG 1453  
Db 99482 TTTAATAGAGACAGATTTCACCATGTTGGCCAACTGCTCTAGAACTCTTAATTCAGG 99423  
QY 1454 TCATCAACCCACCTCAGCTCCCAAGACCTGGATTACAGGCATGAGCCACCGTCCCA 1513  
Db 99422 TCATCCGCTCTCTCAGCTCCCAAGTCTGCGATTACAGGTGTGAGCCACCATGCCCCA 99363  
QY 1514 GCCTGTTTCTCAGATCCTGTATTGTT 1541  
Db 99362 GCCTGCTTTCTTTCTTTCTTATTATT 99335

RESULT 45  
ABD32868/C  
ID ABD32868 standard, DNA, 238417 BP.  
XX  
AC ABD32868;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human cancer-associated genomic DNA HD17-053.  
XX  
KM Human; ds; cancer-associated protein; gene; cytosolic; cancer;  
KW leukaemia; lymphoma; CAP.  
XX  
OS Homo sapiens.  
XX  
PN WO2004074320-A2.  
XX  
PD 02-SEP-2004.  
XX  
PF 17-FEB-2004; 2004WO-US004730.  
XX  
PR 14-FEB-2003; 2003US-00367094.  
PR 14-MAR-2003; 2003US-0038838.  
PR 15-APR-2003; 2003US-00417375.  
PR 13-JUN-2003; 2003US-00461862.  
PR 15-SEP-2003; 2003US-00653431.  
PR 15-DEC-2003; 2003US-00737318.  
XX  
PA (SAGR-) SAGRES DISCOVERY INC.  
XX  
PI Morris DW, Morris DW, Malandro MS;  
XX  
DR WPI, 2004-652914/63.  
XX  
PT New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.  
XX  
PS claim 16; seqid 548; 310bp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above

CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancers, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a human CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 238417 BP; 57141 A; 58273 C; 60612 G; 62391 T; 0 U; 0 Other;

Query Match 10.1%; Score 201.6; DB 13; Length 238417;

Best Local Similarity 82.0%; Pred. No. 5.3e-40;

Matches 269; Conservative 0; Mismatches 54; Indels 5; Gaps 3;

QY 1216 TCTTTTATTTTATTTTATTTTGAACAGAGCTCACTTTGTACACCGCTGAGGTACA 1275  
Db 158308 TCTTTTCTTTTCTTTTCTTTTGAAGAGAGCTGCTGCTGTACCCAGGCTGAGGTACA 158249  
QY 1276 GTGCTGTGATCTGCGCTCACTGCAACCTTGCTCCAGGTTCAAGGATTCCTG-- 1333  
Db 158248 ATGG-TGGAGTCTGCGCTCACTGTAACCTTGCTCCAGGCTCAAGGATTCCTG-- 158190  
QY 1334 TCAGCTTCCGAAATGCTGGGATTTACAGGCGCATGACACCA--TGCTTAATTTTGT 1391  
Db 158189 TCAGCTTCCGAGTGTGAGATTACAGGATGCGCACACACCGGCTAATTCCTGTA 158130  
QY 1392 TTTTATAGTAGAGACAGATTTGCGCATGTTGACCAAGGCTTGCTGTAATCTGTA 1451  
Db 158129 TTTTATAGTAGAGACAGGATTTTCAACCATGTTGCCAGGCTGTCTAAATCTGACCTCA 158070  
QY 1452 GGTATCCACCCACCTCAGCTCCCAAGACCTGGGATTACAGGATGAGCCACCGTGCC 1511  
Db 158069 GGTATCCACCCGCTCAGCTCCCAAGGATGAGGATTACAGGCTGAGCCACCGCAC 158010  
QY 1512 CAGCTGTTTCTCAGATCCTGTATTG 1539  
Db 158009 CAGCCATCTCCAGGAATTTCTATGTG 157982

Search completed: January 22, 2006, 04:09:12

Job time : 817.081 secs

(b)(5) DPP, (b)(7)(C) FOIA

(b)(5) DPP, (b)(7)(C) FOIA



98 198.8 9.9 156651 3 US-09-949-016-17349 Sequence 17349, A  
99 198.6 9.9 601 3 US-09-949-016-165314 Sequence 165314,  
c 100 198.6 9.9 28884 3 US-09-949-016-16401 Sequence 16401, A

## ALIGNMENTS

## RESULT 1

US-09-078-294-3  
; Sequence 3, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; CURRENT FILING DATE: 1998-05-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 80595  
; TYPE: DNA  
; ORGANISM: Nucleotide sequence of HC-contig  
US-09-078-294-3

Query Match 100.0%; Score 2001; DB 3; Length 80595;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATGTTGGATTTCAGAGGTACAGAAAGCTGAAAAACAATCACTAGCCGAGCTTGAGAG 60  
DB 19000 AATGTTGGATTTCAGAGGTACAGAAAGCTGAAAAACAATCACTAGCCGAGCTTGAGAG 19059  
QY 61 TTTGACGGAGACGTGTCAGCCCTGTGTTTTTCCATGACAGCTGAAAAATAGCCCAAC 120  
DB 19060 TTTGACGGAGACGTGTCAGCCCTGTGTTTTTCCATGACAGCTGAAAAATAGCCCAAC 19119  
QY 121 TTCAGTGAAGCTTGTTCCTTCCCTCCTCAAGTTACCAAAATTCAGTTCTCTCAG 180  
DB 19120 TTCAGTGAAGCTTGTTCCTTCCCTCCTCAAGTTACCAAAATTCAGTTCTCTCAG 19179  
QY 181 AAAGCAAAAAATGATTTAGAGGTTAGATTTGGTTCTTTATCTATTCAGAGATTG 240  
DB 19180 AAAGCAAAAAATGATTTAGAGGTTAGATTTGGTTCTTTATCTATTCAGAGATTG 19239  
QY 241 ATATATATGTTCTCCACACAGATGTTCTGCTGTAAACAATCACTCTCTGACACTACTG 300  
DB 19240 ATATATATGTTCTCCACACAGATGTTCTGCTGTAAACAATCACTCTCTGACACTACTG 19299  
QY 301 CATATGACAGAGTGTATCTACCAAGGTAAACACAGAAATGGCTGCCAATTCGAATCCC 360  
DB 19300 CATATGACAGAGTGTATCTACCAAGGTAAACACAGAAATGGCTGCCAATTCGAATCCC 19359  
QY 361 TGAACGAGTGAAGAAATCAGAAATTAATATAGGGGATTCACAGAGCTGGCTAGCGAGT 420  
DB 19360 TGAACGAGTGAAGAAATCAGAAATTAATATAGGGGATTCACAGAGCTGGCTAGCGAGT 19419  
QY 421 TCCAGATGTCAGATATCTTGTGCATCATACGAGGTGCTGTCTTAGCAACTGCTCA 480  
DB 19420 TCCAGATGTCAGATATCTTGTGCATCATACGAGGTGCTGTCTTAGCAACTGCTCA 19479  
QY 481 CTGCTTCATTTCTGCTTGTGCTTTAAATATCTGCTTTTCTCAGCTCAATTGCTTTCTT 540  
DB 19480 CTGCTTCATTTCTGCTTGTGCTTTAAATATCTGCTTTTCTCAGCTCAATTGCTTTCTT 19539  
QY 541 CCTCGGAGTCAAGCTTTCTTTGGGTCAAAAGCAAGATATCTTTAGATCACTGGT 600  
DB 19540 CCTCGGAGTCAAGCTTTCTTTGGGTCAAAAGCAAGATATCTTTAGATCACTGGT 19599  
QY 601 ACTCAAGAGGTACAAAGACATTTGGCATTCCTTCACCTCTTGAAAAACAATTTTA 660

DB 19600 ACTCAAGAGGTACAAAGACATTTGGCATTCCTTCACCTCTTGAAAAACAATTTTA 19659  
QY 661 TGAAGCCAGGTTAGCCATAGTGCCTCTTGAGTTGTTGCTCAGCCAGGCCCAAGCTT 720  
DB 19660 TGAAGCCAGGTTAGCCATAGTGCCTCTTGAGTTGTTGCTCAGCCAGGCCCAAGCTT 19719  
QY 721 TGTGCTTCAACATGAATTAAGAGCTTCAGAAACAGATCCACATTTTCAATGCTTCA 780  
DB 19720 TGTGCTTCAACATGAATTAAGAGCTTCAGAAACAGATCCACATTTTCAATGCTTCA 19779  
QY 781 CCCAATCTGATTAAGAAACAAATTCGCATATCTCAATGACCACTTTTCAAGTGGATAG 840  
DB 19780 CCCAATCTGATTAAGAAACAAATTCGCATATCTCAATGACCACTTTTCAAGTGGATAG 19839  
QY 841 TAGATGCTGAATGGGTTCACAGCAATGGCCCAACCACTTTGCAAAAAAGCTGGAACT 900  
DB 19840 TAGATGCTGAATGGGTTCACAGCAATGGCCCAACCACTTTGCAAAAAAGCTGGAACT 19899  
QY 901 CTGACTGGGACCTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGATATGAACCCC 960  
DB 19900 CTGACTGGGACCTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGATATGAACCCC 19959  
QY 961 GTGATGATATATAGTAAAGGGTTGGCCTTATGTTTCTATTCCTTCAAAACCTGTAG 1020  
DB 19960 GTGATGATATATAGTAAAGGGTTGGCCTTATGTTTCTATTCCTTCAAAACCTGTAG 20019  
QY 1021 AATAGATATGCTTGTTCCTTTAAATAATGCAACAATTCATTTATGATGCTGTAT 1080  
DB 20020 AATAGATATGCTTGTTCCTTTAAATAATGCAACAATTCATTTATGATGCTGTAT 20079  
QY 1081 AGTAACTCAAGATCATCTCCATGAAATATGCTTCAGAACCCATATTAAGAGATTTT 1140  
DB 20080 AGTAACTCAAGATCATCTCCATGAAATATGCTTCAGAACCCATATTAAGAGATTTT 20139  
QY 1141 AGCCATGTGTACAAAGAGGCCATTTCAAGTGTGAATTTGTCAGAGATTTGA 1200  
DB 20140 AGCCATGTGTACAAAGAGGCCATTTCAAGTGTGAATTTGTCAGAGATTTGA 20199  
QY 1201 TTATGTTTTCAGATCTTTTATTTATTTTATTTTGAACAGAGTCTCATTTGTGAC 1260  
DB 20200 TTATGTTTTCAGATCTTTTATTTATTTTATTTTGAACAGAGTCTCATTTGTGAC 20259  
QY 1261 CCAGGCTGAGTACAGTGGCTGTGCTGCTGCTCAACCTGCTCCACAGTTCA 1320  
DB 20260 CCAGGCTGAGTACAGTGGCTGTGCTGCTGCTCAACCTGCTCCACAGTTCA 20319  
QY 1321 AGCGATTCCTGTCAAGCTTCCCGAATAGCTGGGATTAACAGCGCATGACACCATGCC 1380  
DB 20320 AGCGATTCCTGTCAAGCTTCCCGAATAGCTGGGATTAACAGCGCATGACACCATGCC 20379  
QY 1381 TAAATTTGATTTTGAAGAGACAGAGTTGCGCATGTTGACAGGCTTGTGAAC 1440  
DB 20380 TAAATTTGATTTTGAAGAGACAGAGTTGCGCATGTTGACAGGCTTGTGAAC 20439  
QY 1441 TCCTGATCTCAGATGATCCACACCTGACCTCCCAAGCACTGGGATTAACGCGATGA 1500  
DB 20440 TCCTGATCTCAGATGATCCACACCTGACCTCCCAAGCACTGGGATTAACGCGATGA 20499  
QY 1501 GCCACCGTCCAGAGCTGTTTCTAGATCCGTATATTTGTTCTGAAAGCTTCAATTTCA 1560  
DB 20500 GCCACCGTCCAGAGCTGTTTCTAGATCCGTATATTTGTTCTGAAAGCTTCAATTTCA 20559  
QY 1561 TCTTCTTATATCTTTGGAAGATGACCTTAAGTAAAGTTTAAACAATCAATATCTT 1620  
DB 20560 TCTTCTTATATCTTTGGAAGATGACCTTAAGTAAAGTTTAAACAATCAATATCTT 20619  
QY 1621 TGAATAATTCCTGTGCTTTCTTATTCCTAACAAAAATGTTAGTAACTGATGTT 1680  
DB 20620 TGAATAATTCCTGTGCTTTCTTATTCCTAACAAAAATGTTAGTAACTGATGTT 20679  
QY 1681 ATGTTCTTCAATTAATTCATTTCTATCTCAGAAATTTATCTAGTCTAATGTTAT 1740



Db	20680	ATGTTCTTTCAAAATATATCATTTCTCTATCCAGAAATTATCTCAGGCTAAATGTAT	20739
Qy	1741	TGAATAGCTTTCACTTTCTTGTCATCCAGTTTCTGTCTCTATATTTCACTTAAGCTTAAG	1800
Db	20740	TGAATVAGCTTCACTTTTGTGTATCCAGTTTCTGTCTCTATATTTCACTTAAGCTTAAG	20799
Qy	1801	TGGCTATTAGAAATMAAGCTTTGTAAAGATATCTTTCTCCAAATATGCTTAATCTTTTGAC	1860
Db	20800	TGGCTATTAGAAATMAAGCTTTGTAAAGATATCTTTCTCCAAATATGCTTAATCTTTTGAC	20859
Qy	1861	TGCATGCGCAGGACAAACCTGTAACTGTTTGTATCTTCAATACATTTCCACAGAAATGCG	1920
Db	20860	TGCATGCGCAGGACAAACCTGTAACTGTTTGTATCTTCAATACATTTCCACAGAAATGCG	20919
Qy	1921	TGACTCTCTCTTCTCTAAAGCAATGCCCAAGCAGACATGTTAGATAGTATGTACGCA	1980
Db	20920	TGACTCTCTCTTCTCTAAAGCAATGCCCAAGCAGACATGTTAGATAGTATGTACGCA	20979
Qy	1981	ACAGGGACATGGGTGTCATAGC 2001	
Db	20980	ACAGGGACATGGGTGTCATAGC 21000	

RESULT 2  
US-09-078-294-4  
; Sequence 4, Application US/09078294

```

: Patent No. 6265211
: GENERAL INFORMATION:
: APPLICANT: Cho, Kong-Hong Andy
: APPLICANT: Du Sart, Desiree
: APPLICANT: Cancilla, Michael R.
: TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
: FILE REFERENCE: Davies Col
: CURRENT APPLICATION NUMBER: US/09/078,294
: CURRENT FILING DATE: 1998-05-13
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 80246
: TYPE: DNA
: ORGANISM: Nucleotide sequence of NC-contig
: US-09-078-294-4

```

Query Match	98.2%	Score 1964.6	DB 3	Length 80246
Best Local Similarity	99.5%	Pred. No. 0		
Matches 1992, Conservative	0	Mismatches	9	Indels 2; Gaps 2;

QY	1	AAAGTTGGAAATTC	CAAGGTAA	CAAGAA	AGCTGG	AAAA	CAAA	CTAC	ATGG	CCG	CA	AGTCTG	AGG	60
Db	18736	AAATGTTTGAAATTC	CAAGGTAA	CAAGAA	AGCTGG	AAAA	CAAA	CTAC	TG	CCG	AGTCTG	AGG	18795	
QY	61	TTTCAGCGGAG	ACTGGT	GCAG	CCCTTG	TGTTTTC	CA	CTGA	CAG	CTAA	AAATG	AGCC	120	
Db	18796	TTTCAGCGGAG	ACTGGT	GCAG	CCCTTG	TGTTTTC	CA	CTGA	CAG	CTAA	AAATG	AGCC	18855	
QY	121	TTTCAGTGA	AGCTTGT	TCTT	CCCTCCT	CTCA	AGTTAC	CC	CAATTT	CTCAG	TTCT	CTCA	180	
Db	18856	TTTCAGTGA	AGCTTGT	TCTT	CCCTCCT	CTCA	AGTTAC	CC	CAATTT	CTCAG	TTCT	CTCA	18915	
QY	181	AAAGCCAAAA	AATGAATTT	GAGGGTT	AGATTTG	GGTTC	TTTAT	CTATTA	CA	AGATG			240	
Db	18916	AAAGCCAAAA	AATGAATTT	GAGGGTT	AGATTTG	GGTTC	TTTAT	CTATTA	CA	AGATG			18975	
QY	241	ATATATAT	GTTCCT	CCAC	CAAGATG	TTCG	CTTGTA	CA	ATAT	CTCA	CTTCT	CTGAC	300	
Db	18976	ATATATAT	GTTCCT	CCAC	CAAGATG	TTCG	CTTGTA	CA	ATAT	CTCA	CTTCT	CTGAC	19035	
QY	301	CATATG	CAGAGTGT	TACT	CTAC	CAAGGTAA	CA	CAGAA	TGGCTG	CCCA	ATTC	CCAAT	360	
Db	19036	CATATG	CAGAGTGT	TACT	CTAC	CAAGGTAA	CA	CAGAA	TGGCTG	CCCA	ATTC	CCAAT	19095	
QY	361	TGAATCTGA	GTGAG	GAATCAG	AATTA	TATATAG	GGGATTC	AC	AGACTG	GGCT	CTAC	GGATG	420	

Db	19096	TGAATGAGTGAAGAGAAATTCAGAAATTATTAATAGGGATTCACACAGAGCTGGCTACGAGATG	191535
OY	421	TGCCAGTGGTTCAGATACCTTGTGCTCATACGAGAGGTGGCTGCTCTTAAGCACTGGTCA	480
Db	19156	TGCCAGTGGTTCAGATACCTTGTGCTCATACGAGAGGTGGCTGCTCTTAAGCACTGGTCA	192151
OY	481	CTGGTTCAATTTCCCTGGCTTGGTCTTTAAATACGTGCTTTTCTCAGCTCAATTGGCTTTCTT	540
Db	19216	CTGGTTCAATTTCCCTGGCTTGGTCTTTAAATACGTGCTTTTCTCAGCTCAATTGGCTTTCTT	192759
OY	541	CCCTCTGGCAGTCAAGCTTTCTTTGGGGTCAAAACAGCAAAATGATTTCTTTAATATCACCTGGT	600
Db	19276	CCCTCTGGCAGTCAAGCTTTCTTTGGGGTCAAAACAGCAAAATGATTTCTTTAATATCACCTGGT	193353
OY	601	ACTCAAAAGAGCTACACAGACATTTGGGCAATTCACCTTCACCTCTTTGGAAAAACAATTTTA	660
Db	19336	ACTCAAAAGAGCTACACAGACATTTGGGCAATTCACCTTCACCTCTTTGGAAAAACAATTTTA	193959
OY	661	TGGAAAGCCAAAGTTGCCATATGTGCTCTTGAAGTTGTTTGTCTCAGCCAAAGGCCAAAGCTT	720
Db	19396	TGGAAAGCCAAAGTTGCCATATGTGCTCTTGAAGTTGTTTGTCTCAGCCAAAGGCCAAAGCTT	194555
OY	721	TGTGCTTCAAAACATGAAATTTAGAGAGCTTCAGAAACAAGATCCACATTTTCAATGGCTTCA	780
Db	19456	TGTGCTTCAAAACATGAAATTTAGAGAGCTTCAGAAACAAGATCCACATTTTCAATGGCTTCA	195151
OY	781	CCCAACTGATATTAAGAACAATTTGCCATATCTTCATATGACCACTTTT-TCAGTGGGAGT	839
Db	19516	CCCAACTGATATTAAGAACAATTTGCCATATCTTCATATGACCACTTTTCTCAGTGGGAGT	195755
OY	840	GTAATGCTGGAAATGGGTCAAGACATTTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC	899
Db	19576	GTAATGCTGGAAATGGGTCAAGACATTTGCCCAACCAAACTTTGCAAAAAAGGCTGGAAGC	196355
OY	900	TTGCACTGGGGAGCCCTTAAATATGCAAAAAGTTGATAGGCTCTTATGCAAGAAATATGAATCC	959
Db	19636	TTGCACTGGGGAGCCCTTAAATATGCAAAAAGTTGATAGGCTCTTATGCAAGAAATATGAATCC	196955
OY	960	CGTGTATGATATATAGCTTAAAGGGTGGGCTTTATGTTTCTATTCCCTTCACAAACCTGGTA	1019
Db	19696	CGTGTATGATATATAGCTTAAAGGGTGGGCTTTATGTTTCTATTCCCTTCACAAACCTGGTA	197555
OY	1020	GAATAGATATGCTGTTTCCCTTTAAAAAATGTCAACAAATTCATTTATGATGCTGTGTA	1079
Db	19756	GAATAGATATGCTGTTTCCCTTTAAAAAATGTCAACAAATTCATTTATGATGCTGTGTA	198151
OY	1080	TATGTAATCTCAAGATCATGTCTCATGAAAAATGCTTGAACCCCATATATAGAGAAATTTT	1139
Db	19816	TATGTAATCTCAAGATCATGTCTCATGAAAAATGCTTGAACCCCATATATAGAGAAATTTT	198755
OY	1140	TAGCCATGTGTGACAAAAGAGAGGCAATTCAGTGTGAAATTTGTTACAGAGAAAGATTTTG	1199
Db	19876	TAGCCATGTGTGACAAAAGAGAGGCAATTCAGTGTGAAATTTGTTACAGAGAAAGATTTTG	199355
OY	1200	ATTATGTTTTCTCAGATCTTTTATTTTATTTTTTTGAACAGAGTCTCACTTTGTCA	1259
Db	19936	ATTATGTTTTCTCAGATCTTTTATTTTATTTTTTTGAACAGAGTCTCACTTTGTCA	199955
OY	1260	CCCAAGGCTGGAATACAGTGGCTGTGTCTTCGGCTCATCTGCAACCTCTGCTTCCAGGTTTC	1319
Db	19996	CCCAAGGCTGGAATACAGTGGCTGTGTCTTCGGCTCATCTGCAACCTTTGCTTCCAGGTTTC	200555
OY	1320	AAGGATTTCTCTGTGAGCTTCCCGAAATATGCTGGAGTTACAGAGGCAATGACACACATGTC	1379
Db	20056	AAGGATTTCTCTGTGAGCTTCCCGAAATATGCTGGAGTTACAGAGGCAATGACACACATGTC	201155
OY	1380	CTAATTTTTGTATTTTTTATGAGACACAGATTTCCGCATGTTGACCAAGGCTTGCCTTGA	1439
Db	20116	CTAATTTTTGTATTTTTTATGAGACACAGATTTCCGCATGTTGACCAAGGCTTGCCTTGA	201755
OY	1440	CTCTGTACTCAGGTATCCACCCATCTGAGCTCCCAAGACATGGGATTTACAGGCAATG	1499
Db	20176	CTCTGTACTCAGGTATCCACCCATCTGAGCTCCCAAGACATGGGATTTACAGGCAATG	202355

QY	960	CGGTATGAGATATGAGCTCAAAAGGGTTGGCCCTTTATGTTTCTATCTCTCAAAACCTGGTA	1019
Db	19696	CGGTATGAGATATGCTCAAAAGGGTTGGCCCTTTATGTTTCTATCTCTCAAAACCTGGTA	19755
QY	1020	GAATAGATATGCTGTTTCCCTTTAAAAAATGTCAAATTCGATTATGATGCTGTGTA	1079
Db	19756	GAATAGATATGCTGTTTCCCTTTAAAAAATGTCAAATTCGATTATGATGCTGTGTA	19815
QY	1080	TAGTAATCTCAGATCATGCTCCATGATAAATGTTTCAGAAACCCAAATATTAAGAGATTTT	1139
Db	19816	TAGTAATCTCAGATCATGCTCCATGATAAATGTTTCAGAAACCCAAATATTAAGAGATTTT	19875
QY	1140	TAGCCATGTGTGACAAAAAGAGAGCCATTTCAGTGTGAAATGTTCAGAGAAATATTG	1199
Db	19876	TAGCCATGTGTGACAAAAAGAGAGCCATTTCAGTGTGAAATGTTCAGAGAAATATTG	19935
QY	1200	ATTATGTTTTCTCAGATCTTTTATTTTATTTTATTTTTCAAAACAGTCTCATCTTGTA	1259
Db	19936	ATTATGTTTTCTCAGATCTTTTATTTTATTTTATTTTTCAAAACAGTCTCATCTTGTA	19995
QY	1260	CCCAGGCTGAGATACAGTGGCTGTGCTTCGCTCACTGCAACCTTCGCTCCAGGTTTC	1319
Db	19996	CCCAGGCTGAGATACAGTGGCTGTGCTTCGCTCACTGCAACCTTCGCTCCAGGTTTC	20055
QY	1320	AAGCGATTCTTCCTGTCAAGTTCCCGAAATAGCTGGATTACAGGCGCATGCAACCAATGC	1379
Db	20056	AAGCGATTCTTCCTGTCAAGTTCCCGAAATAGCTGGATTACAGGCGCATGCAACCAATGC	20115
QY	1380	CTAATTTTGTATTTTATGATAGACAGAGTTTCGCCATGTTGACCAAGCTTGCCTTGA	1439
Db	20116	CTAATTTTGTATTTTATGATAGACAGAGTTTCGCCATGTTGACCAAGCTTGCCTTGA	20175
QY	1440	CTCTGATCTCAGGTATCCACCACCTCAGCTCCCAAAAGACATGGATTACAGGCATG	1499
Db	20176	CTCTGATCTCAGGTATCCACCACCTCAGCTCCCAAAAGACATGGATTACAGGCATG	20235

OY	840	GTAGATCTGGAAATGGGTCA	CAGCATTGGCCAA	CCAAACCTTTGC	CAAAAAAGGCTGAAGC	899
Db	19576	GTAGATCTGGAAATGGGTCA	CAGCATTGGCCAA	CCAAACCTTTGC	CAAAAAAGGCTGAAGC	19635
OY	900	TTTCGACTGGGGACCCCTTAA	ATATGC	AAAAGTTGATAGGCTCTT	CATGCAGAAATATGAACC	959
Db	19636	TTTCGACTGGGGACCCCTTAA	ATATGC	AAAAGTTGATAGGCTCTT	CATGCAGAAATATGAACC	19695
OY	960	CGTGTATGATATATAGCTTAA	AGGGTGGGCTTTATGTTT	CTATTCTCTTCA	CAAAACCTGGTA	1019
Db	19696	CGTGTATGATATATAGCTTAA	AGGGTGGGCTTTATGTTT	CTATTCTCTTCA	CAAAACCTGGTA	19755
OY	1020	GAATAGATATGCTTGTTTCC	CTTTAAAAAATGTCA	CAATTCGATTATGATGCTGTGTA		1079
Db	19756	GAATAGATATGCTTGTTTCC	CTTTAAAAAATGTCA	CAATTCGATTATGATGCTGTGTA		19815
OY	1080	TAGTAATCTCAGATCATAGCT	CCATGATAAAATGGTTAGAA	ACCCAATATATAGAGATTTT		1139
Db	19816	TAGTAATCTCAGATCATAGCT	CCATGATAAAATGGTTAGAA	ACCCAATATATAGAGATTTT		19875
OY	1140	TAGCCATGTGTGACAA	AAAGAGAGGCCATTTCAGTGTGA	ATTTGTTCAGAGAAATATTG		1199
Db	19876	TAGCCATGTGTGACAA	AAAGAGAGGCCATTTCAGTGTGA	ATTTGTTCAGAGAAATATTG		19935
OY	1200	ATTATGTTTTCTCAGATCT	TTTTTATTTTTTATTTTTT	TGAAAACAGAGTCTACTTTGTA		1259
Db	19936	ATTATGTTTTCTCAGATCT	TTTTTATTTTTTATTTTTT	TGAAAACAGAGTCTACTTTGTA		19995
OY	1260	CCGACGCTGAGATACAGT	GGCTGTGGTCTCGGCTCA	CTGCAACCTTGCTCCAGGTTTC		1319
Db	19996	CCGACGCTGAGATACAGT	GGCTGTGGTCTCGGCTCA	CTGCAACCTTGCTCCAGGTTTC		20055
OY	1320	AAGCGATTCTTCCTGTCA	GGCTTCCGAAATAGCTGGGATTA	CAGGCGCATGCAACCAATGC		1379
Db	20056	AAGCGATTCTTCCTGTCA	GGCTTCCGAAATAGCTGGGATTA	CAGGCGCATGCAACCAATGC		20115
OY	1380	CTAATTTTTGTATTTTTT	AGTAGAGACAGAGTTT	CGCCATGTTGAC	CAGGCTTGCCCTTGA	1439
Db	20116	CTAATTTTTGTATTTTTT	AGTAGAGACAGAGTTT	CGCCATGTTGAC	CAGGCTTGCCCTTGA	20175
OY	1440	CTCTGTGACTCAGGTGAT	CCACCACTCAGGCTCC	CAAAAGACATGGGATTA	CAGGCAATG	1499
Db	20176	CTCTGTGACTCAGGTGAT	CCACCACTCAGGCTCC	CAAAAGACATGGGATTA	CAGGCAATG	20235

QY	721	TGTCCTTCAACATGGAATTTAGAGACCTTCAGAACAGATCCACATTTTCAATGGCTTCA	780
Db	19456	TGTCTTCAAAACATGAATTTAGAGACCTTCAGAACAGATCCACATTTTCAATGGCTTCA	19515
QY	781	CCCACTCGGATTAAGAAACAATTGCCATATCTCAATGACCACTTTT-TCAGTGGGATG	839
Db	19516	CCCACTCGGATTAAGAAACAATTGCCATATCTCAATGACCACTTTTTCAGTGGGATG	19575
QY	840	GTAGATCTGGAAATGGGTCCACGATTTGGCCAAACAACTTTGCCAAAAAGGCTGGAAAGC	899
Db	19576	GTAGATCTGGAAATGGGTCCACGATTTGGCCAAACAACTTTGCCAAAAAGGCTGGAAAGC	19635
QY	900	TCTGACTGGGAGACCCCTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGAAATGAACCC	959
Db	19636	TCTGACTGGGAGACCCCTTAATATGCAAAAGTTGATAGGCTCTTCATGACAGAAATGAACCC	19695
QY	960	CGTGTATGATPATAGCTAAAGGGGTGGCCTTATATGTTTCTATTCCTTCCAAACCTGGTA	1019
Db	19696	CGTGTATGATPATAGCTAAAGGGGTGGCCTTATATGTTTCTATTCCTTCCAAACCTGGTA	19755
QY	1020	GAATGATATGCTTTGTTCCCTTTAAAAAATGTCAACAAATTCGATTTATGATGCTGTGTA	1079
Db	19756	GAATGATATGCTTTGTTCCCTTTAAAAAATGTCAACAAATTCGATTTATGATGCTGTGTA	19815
QY	1080	TAGTAATCTCAGATCATGCTCCATGAAAAATGCTCAGAAACCAATATTAAGAGATTTT	1139
Db	19816	TAGTAATCTCAGATCATGCTCCATGAAAAATGCTCAGAAACCAATATTAAGAGATTTT	19875
QY	1140	TAGCATGTGTGACAAAGAAGAGGCCATTTAGTGTGTGAATTGTTCAGAGAAGTATTTG	1199
Db	19876	TAGCATGTGTGACAAAGAAGAGGCCATTTAGTGTGTGAATTGTTCAGAGAAGTATTTG	19935
QY	1200	ATTATGTTTTCTCAGATCTTTTATTTTATTTTGTGAAAACAGAGCTCACTTTGTCA	1259
Db	19936	ATTATGTTTTCTCAGATCTTTTATTTTATTTTGTGAAAACAGAGCTCACTTTGTCA	19995
QY	1260	CCCAAGCTGGAATACAGTGGCTGTGGTCTCCGCTCATCTGCAACCTCTGCTCCCAAGTTTC	1319
Db	19996	CCCAAGCTGGAATACAGTGGCTGTGGTCTCCGCTCATCTGCAACCTCTGCTCCCAAGTTTC	20055
QY	1320	AAGGATTTCTCTGTAGGCTTCCCGAATATGCTGGGATTCAGAGGCGATTCACACATATGC	1379
Db	20056	AAGGATTTCTCTGTAGGCTTCCCGAATATGCTGGGATTCAGAGGCGATTCACACATATGC	20115
QY	1380	CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCCGCATGTTGACCAAGGCTTGCTTGAA	1439
Db	20116	CTAATTTTTGTATTTTTAGTAGAGACAGAGTTTCCGCATGTTGACCAAGGCTTGCTTGAA	20175
QY	1440	CTCTGATCTCAGGTATCCACCACTGAGCTCCCAAGACACTGGGATTCACAGGCAATG	1499
Db	20176	CTCTGATCTCAGGTATCCACCACTGAGCTCCCAAGACACTGGGATTCACAGGCAATG	20235

Db	19276	CCCTCTGGCAGTCACTGTTTCTTTGGGTCAAACAGCAAAATGATCTTTAGATATCACTGGT	19335
QY	601	ACTCAAAAGGAGCTACAAAGACATTGGGAGTCACCTTCCACTCTCTTGGAAAAACAATTTTA	660
Db	19336	ACTCAAAAGGAGCTACAAAGACATTGGGAGTCACCTTCCACTCTCTTGGAAAAACAATTTTA	19395
QY	661	TGGAAAGCAAGGTTGCCATGTGCTCCCTTGAGGTTGTGGCTCAGCAAGGCCCAAGCTT	720
Db	19396	TGGAAAGCAAGGTTGCCATGTGCTCCCTTGAGGTTGTGGCTCAGCAAGGCCCAAGCTT	19455
QY	721	TGTGCTTCAAAACATGAAATTAGAGAGCTTCAGAAACAAGATCCACATTTTCAATGGCCTCA	780
Db	19456	TGTGCTTCAAAACATGAAATTAGAGAGCTTCAGAAACAAGATCCACATTTTCAATGGCCTCA	19515
QY	781	CCCAACTGATPAAAAGAACATTTGCCATATCTCATGACCACTTTT-TGAGTGGGATG	839
Db	19516	CCCAACTGATPAAAAGAACATTTGCCATATCTCATGACCACTTTTCTCAGTGGGATG	19575
QY	840	GTCAGATCTCTGGAAATGGGTCAACAGATTGGCCCAACCAACTTTGGCAAAAAAGGCTGGAAGC	899
Db	19576	GTCAGATCTCTGGAAATGGGTCAACAGATTGGCCCAACCAACTTTGGCAAAAAAGGCTGGAAGC	19635
QY	900	TTCGACTGGGGAGCCCTPAAATATGCAAAAGTGTGATAGGCTTCATCAGCAGAAATATGAACC	959
Db	19636	TTCGACTGGGGAGCCCTPAAATATGCAAAAGTGTGATAGGCTTCATCAGAAATATGAACC	19695
QY	960	CGTGATGATATATAGCTPAAAGGGTTGGCCTTTATGTGTTCTATTCCTTCACAAACCTGGTA	1019
Db	19696	CGTGATGATATATAGCTPAAAGGGTTGGCCTTTATGTGTTCTATTCCTTCACAAACCTGGTA	19755
QY	1020	GAATAGATATGCTTGTGTTCCCTTTAAAAAATGTCAACAATTCGATTATGATGCTGTGTA	1079
Db	19756	GAATAGATATGCTTGTGTTCCCTTTAAAAAATGTCAACAATTCGATTATGATGCTGTGTA	19815
QY	1080	TAGTAATCTCAACATCATAGCTCATGATAAATGTTCCAGAACCCCAATATATAGAGATTTTT	1139
Db	19816	TAGTAATCTCAACATCATAGCTCATGATAAATGTTCCAGAACCCCAATATATAGAGATTTTT	19875
QY	1140	TAGCATGTGTGACAAAAAGAGAGGCCATTTCAGTGTGGAATTGTTCAGAGAAATATTTG	1199
Db	19876	TAGCATGTGTGACAAAAAGAGAGGCCATTTCAGTGTGGAATTGTTCAGAGAAATATTTG	19935
QY	1200	ATTATGTTTTCTCAGATCTTTTATTTTATTTTTTTGAAAACAAGTCTCACTTTGTCA	1259
Db	19936	ATTATGTTTTCTCAGATCTTTTATTTTATTTTTTTGAAAACAAGTCTCACTTTGTCA	19995
QY	1260	CCCAAGGCTGAGATACAGTGGCTGTGGTCTCGGCTCATCTGCAACCTCTGCTTCCAGGTTTC	1319
Db	19996	CCCAAGGCTGAGATACAGTGGCTGTGGTCTCGGCTCATCTGCAACCTTTGCTTCCAGGTTTC	20055
QY	1320	AAGGAGATTTCTCGTGAAGCTTCCCGAATATGCTGGGATTCAGAGGCGATATCCACACATATG	1379
Db	20056	AAGGAGATTTCTCGTGAAGCTTCCCGAATATGCTGGGATTCAGAGGCGATATCCACACATATG	20115
QY	1380	CTAATTTTTGTATTTTTTATGTAGAGACAGAGTTTCGCGCATGTTGACCAAGGCTTGCTTGAA	1439
Db	20116	CTAATTTTTGTATTTTTTATGTAGAGACAGAGTTTCGCGCATGTTGACCAAGGCTTGCTTGAA	20175
QY	1440	CTCTGTACTTCAGGTATCCACCCACTGAGCTCCCAAGACATGGGATTTACAGGCAAG	1499
Db	20176	CTCTGTACTTCAGGTATCCACCCACTGAGCTCCCAAGACATGGGATTTACAGGCAAG	20235

Db	19156	TGCCAGAGGTACAGTACTTGTGCTCATCAACGAGGTGCTGCTGTGCAACTGTCTCA	19215
QY	481	CTGCTTAATTCGCGCTTGCTGCTTTAAATACGCGCTTTCTAGCTCAATGGCTTCTT	540
Db	19216	CTGCTTAATTCGCGCTTGCTGCTTTAAATACGCGCTTTCTAGCTCAATGGCTTCTT	19275
QY	541	CCCTCTGCGATCACGTTTCTTTGGGTCAAAACGCAAAATGATCTTTAGAAACACCTGCT	600
Db	19276	CCCTCTGCGATCACGTTTCTTTGGGTCAAAACGCAAAATGATCTTTAGAAACACCTGCT	19335
QY	601	ACTCAAAAGAGCTACAAAGACATTTGGGCACTTCCACTCTCTTGAAAAACAATTTTA	660
Db	19336	ACTCAAAAGAGCTACAAAGACATTTGGGCACTTCCACTCTCTTGAAAAACAATTTTA	19395
QY	661	TGGAAGCCAAGTGTGCATAGTCCCTCTTGAGGTTGTTTGTCAAGCAAGGCCAAGCTT	720
Db	19396	TGGAAGCCAAGTGTGCATAGTCCCTCTTGAGGTTGTTTGTCAAGCAAGGCCAAGCTT	19455
QY	721	TGTGCTTCAAAATGAATTTAGAGACTTCAGAACAGATCCAATTTCAATGGCTCA	780
Db	19456	TGTGCTTCAAAATGAATTTAGAGACTTCAGAACAGATCCAATTTCAATGGCTCA	19515
QY	781	CCCAACTGGAATPAAAAAGAACATTGCCATPCTCAATGACACACTTTT-TCAGGTGGGATG	839
Db	19516	CCCAACTGGAATPAAAAAGAACATTGCCATPCTCAATGACACACTTTTCTCAGTGGGATG	19575
QY	840	GTAATCTCTGGAATGGGTCAACAGCATTTGCCAACCAACTTTTGCAAAAAAGGCTGGAAGC	899
Db	19576	GTAATCTCTGGAATGGGTCAACAGCATTTGCCAACCAACTTTTGCAAAAAAGGCTGGAAGC	19635
QY	900	TCTGACCTGGGGACCCCTPAAATATGCAAAAGTTGATAGGCTCTTGACAGAAATATGAACC	959
Db	19636	TCTGACCTGGGGACCCCTPAAATATGCAAAAGTTGATAGGCTCTTGACAGAAATATGAACC	19695
QY	960	CGTGTATGATATATAGCTPAAAGGGTTGGCCTTTATGTTTCTATTCCCTTCAAAACCTGGTA	1019
Db	19696	CGTGTATGATATATAGCTPAAAGGGTTGGCCTTTATGTTTCTATTCCCTTCAAAACCTGGTA	19755
QY	1020	GAATAGATATGCTTGTGTTCCCTTTAAAAAAATGTCAACAAATGCAATTATGATGCTGTGTA	1079
Db	19756	GAATAGATATGCTTGTGTTCCCTTTAAAAAAATGTCAACAAATGCAATTATGATGCTGTGTA	19815
QY	1080	TAGTAATCTCAGATCAATGCTCCATGATAAAATGTTGAGAAACCCAAATATATAGAGATTTT	1139
Db	19816	TAGTAATCTCAGATCAATGCTCCATGATAAAATGTTGAGAAACCCAAATATATAGAGATTTT	19875
QY	1140	TAGCCATGTGTGACAAAAAGAGAGGCCATTTCAAGTGTGAATGTGTCAGAGAAATATTTG	1199
Db	19876	TAGCCATGTGTGACAAAAAGAGAGGCCATTTCAAGTGTGAATGTGTCAGAGAAATATTTG	19935
QY	1200	ATTATGTTTTCTCAGATCTTTTATTTTATTTTATTTTATTTTGAACAAGCTCACCTTGTA	1259
Db	19936	ATTATGTTTTCTCAGATCTTTTATTTTATTTTATTTTATTTTGAACAAGCTCACCTTGTA	19995
QY	1260	CCCAAGCTGAGATACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTTC	1319
Db	19996	CCCAAGCTGAGATACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTTC	20055
QY	1320	AAGCGATTTCTCTGTCAAGCTTCCGAAATAGCTGGGATTAACAGCGCATGCAACCAATGC	1379
Db	20056	AAGCGATTTCTCTGTCAAGCTTCCGAAATAGCTGGGATTAACAGCGCATGCAACCAATGC	20115
QY	1380	CTAATTTTTGATTTTTTATGAGAGACAGAGTTTCCGCATGTTGACCAAGCTTGCCCTTGA	1439
Db	20116	CTAATTTTTGATTTTTTATGAGAGACAGAGTTTCCGCATGTTGACCAAGCTTGCCCTTGA	20175
QY	1440	CTCTGATCTCAGGTATCAACCAACTCAGCTCCCAAAAGACATGGGATTAACAGGCATG	1499
Db	20176	CTCTGATCTCAGGTATCAACCAACTCAGCTCCCAAAAGACATGGGATTAACAGGCATG	20235

QY	1500	AGCCACCGTGGCCAGCCGCTTTTCTCAGATCCCTGTA-TTTGTTTCTGAAGCTTCATTTG	1558
Db	20236	AGCCACCGTGGCCAGCCGCTTTTCTCAGATCCCTGTAATTTGTTTCTGAAGCTTCATTTG	20295
QY	1559	TATGCTCTTATTCATTTTGAAGTAGTACACCTAAGTAAGGTTTTCATCAATCAATATC	1618
Db	20296	TATCTCTTATTCATTTTGAAGTAGTACACCTAAGTAAGGTTTTCATCAATCAATATC	20355
QY	1619	TTTGGAAATTTCCCTGTTCTCTTTCTTATTCCTACAAAATAATGTCAGTATAGCTGATG	1678
Db	20356	TTTGGAAATTTCCCTGTTCTCTTTCTTATTCCTACAAAATAATGTCAGTATAGCTGATG	20415
QY	1679	TTATGTTTCTTTCAAATTAATTCATTTCTATCTCAGAAATTAATCTCATGCTTAATGTT	1738
Db	20416	TTATGTTTCTTTCAAATTAATTCATTTCTATCTCAGAAATTAATCTCATGCTTAATGTT	20475
QY	1739	ATTGAATAGTCTTCACTCTTGTGTATCCAGATTTCTGCTCTCTTATTTCACTCTTAAGCTTA	1798
Db	20476	ATTGAATAGTCTTCACTCTTGTGTATCCAGATTTCTGCTCTCTTATTTCACTCTTAAGCTTA	20535
QY	1799	AGTGGCTATTGAATTAAGAAGCTGTATACAGATTTCTTCTCCAAATATGCTTATCTTTTG	1858
Db	20536	ATTGGCTATTGAATTAAGAAGCTGTATACAGATTTCTTCTCCAAATATGCTTATCTTTTG	20595
QY	1859	ACTGCATCCAGTGTACAAACTGTTAAGTTTGAATTTCTTCATTAACATTCACAGAAACAT	1918
Db	20596	ACTGCATCCAGTGTACAAACTGTTAAGTTTGAATTTCTTCATTAACATTCACAGAAACAT	20655
QY	1919	GCTGACTCCTCTCTTCTCTGAAGAAGCATGCCAAGACAGCATTTGTTAGATAGTATGACG	1978
Db	20656	GCTGACTCCTCTCTCTCTGAAGAAGCATGCCAAGACAGCATTTGTTAGATAGTATGACG	20715
QY	1979	CAACGAGGACATGGGTGCATAGC 2001	
Db	20716	CAACGAGGACATGGGTGCATAGC 20738	

```

      RESULT 3
      US-09-949-002-600/c
      ; Sequence 600, Application US/09949002
      ; Patent No. 6900016
      ; GENERAL INFORMATION:
      ; APPLICANT: VENTER, J. Craig et al.
      ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
      ; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
      ; TITLE OF INVENTION: AND USES THEREOF
      ; FILE REFERENCE: CL000790
      ; CURRENT APPLICATION NUMBER: US/09/949,002
      ; CURRENT FILING DATE: 2000-01-28
      ; PRIOR APPLICATION NUMBER: 60/231,401
      ; PRIOR FILING DATE: 2000-09-08
      ; NUMBER OF SEQ ID NOS: 10823
      ; SOFTWARE: FastSeq for Windows Version 4.0
      ; SEQ ID NO 600
      ; LENGTH: 72048
      ; TYPE: DNA
      ; ORGANISM: Human
      ; FEATURE:
      ; NAME/KEY: misc_feature
      ; LOCATION: (1)...(72048)
      ; OTHER INFORMATION: n = A,T,C or G
      US-09-949-002-600

```

Query	Match	Similarity	Score	DB	Length
Best	Local	Similarity	74.0%	Pred.	No. 8.9e-47;
Matches	311;	Conservative	0;	Mismatches	104;
				Indels	5;
				Gaps	3;
QY	1208	TTCTCAGATCTTTTATTTTATTTTGAACAAGATCTCATTTGTACCCAGGCT	1267		
DB	64752	TACTTGATTTCTTTTCTTTTCTTTTATTTTGAAGCAGAGCTCTCTCTGTGCACCCAGGCT	64693		
QY	1258	GGAGTACAGTGGCTGTGGCTCTGGCTCACTCAACCTTGCTCCAGTTCAAGGATT	1327		

Db	64692	GGAGTGGAGTGGC--ACGATCTCAGCTCACTGGCACTCTGCTCTCAGGTTCAAGCATTT	64634
Qy	1328	CTCTGTG--TCAGCTTCCCGAATTCGTGGATTACAGGCGCATGCACACA--TGCCTTAA	1383
Db	64693	CTCTGCTTACGCTCCCGCGATGAGCTGGGATTACAGGTGCCAGGCACACACACTTGGCTAC	64574
Qy	1384	TTTTTTGATTTTATAGTAGAGACAGATTTGGCATGTTGACAGAGCTTGCCCTGAACTCC	1443
Db	64573	TTTTTGTATTTTATAGTAGGAGACGGGTTTGGCCATGTTGTTAGAGCTGTCTTAACTCC	64514
Qy	1444	TGACTTCAGGTGATTCACACCACCTCAGCTCCCAAGCAGCTGGGATTACAGGCATGAGCC	1503
Db	64513	TGACTTCAGGTGATTCACACCACCTTGGCTCCCAAGTCTTGGGATTACAGGCATGAACC	64454
Qy	1504	ACCGTGCACAGCTGTTTCTCAGATCTGTAATTTGTTCTGAAGCTTCATTTCTAATCT	1563
Db	64453	ACCGTGCACCGCATGCTGTGATTTTCTTTGTAAATTTATTTCTTAAGTATTTTAACTCT	64394
Qy	1564	TCTATATCATTTTGGAGTAGTAGTACACTAGTAGAGTTTTTAAACATCAAAATATCTTTGG	1623
Db	64393	TTTTTAATCTTATGTGAATGGAGTTTTTTAAATTTATGTTTTTGGATTTGTTATTTCTAG	64334

```

RESULT 4
US-09-949-002-684/C
; Sequence 684, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 72048
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(72048)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-684

```

Query Match	10.7%;	Score 213.6;	DB 3;	Length 72048;
Best Local Similarity	74.0%;	Pred. No. 8.9e-04;		
Matches 311;	Conservative 0;	Mismatches 104;	Indels 5;	Gaps 3;
Qy 1208	TTTCAGATCTTTTATATTTTATTTTATTTTGTGAACAGAGCTCAGCTTGTACCCAGGCT			1267
Db 64752	TACTGTGATTTTCTTTTCTTTTCTTTTCTTTTGTGACAGAGTCTCTCTGTACACCAAGGCT			646933
Qy 1268	GGAGTACAGTGGCTGTGGTCTCGGCTCAGTCACCACTTGCCCTCCAGAGTTCAAGGATT			1327
Db 64652	GGAGTGCAGTGGC-ACGATCTCAGCTACAGCAACTTGCCCTCAGGTTCAAGGATT			646334
Qy 1338	CTCTCG--TCAGCTTCCCGAATAGCTGGATTACAGGCGCATGCACCACTA--TGCCTTAA			1383
Db 64633	CTCTGACCTTAGCTCCCGAGTAGCTGGGATTACAGGTGCACGCCACCACTGCTAC			645774
Qy 1384	TTTTTGTATTTTATGTATGACAGACAGATTTGGCCATGTTCACACAGGCTTTCCTTGAACCTCC			1443
Db 64573	TTTTTGTATTTTATGTATGAGGACAGAGGTTTGGCCATGTGTGCTCAGGCTGTGTTTGAACCTCC			645154
Qy 1444	TGACCTTACAGTGATCCACCCACCTCAGCTCCCAAGAACATGTGGGATTACAGGCATAGCC			1503
Db 64513	TGACCTCAGTGATCCACCCACTTGGCTCTCCCAAGATGTGGGATTACAGGCATTAACC			64455

```

OY      1504  ACCGTCGCCAGCCTGTTTCTCAGATCCCTGATATTTGTCGAAAGCCTCATTTCTATCT 1563
DB      64453  ACCGTCGCCAGCCTGTTTCTCAGATCTTGTATTTCTTTGTTAAATTTATTCCTTAAGTATTTTATTTCT 64394

OY      1564  TCTTATTCATTTTGGAAAGTAGTACACCTAAGTAAAGCTTTTAAACAATCAATATCTTTG 1623
DB      64393  TTTTATCTATTTGTAATAGGATTTTAAATAATATGTTTTGATGTTTATTTCTAG 64334

RESULT 5
US-09-949-016-17468
; Sequence 17468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17468
; LENGTH: 88906
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1) ..(88906)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17468

Query Match      10.4%; Score 208; DB 3; Length 88906;
Best Local Similarity 78.0%; Pred. No. 3,2e-45;
Matches 276; Conservative 0; Mismatches 75; Indels 3; Gaps 2;

OY      1216  TCTTTTATTTTATTTTATTTTGTGAAACAGAGTCTCATTTGTACCCAGCTGAGATACA 1275
DB      36054  TTTTATTTTATTTTATTTTATTTTGTGAATGAGTCTGTCTGTACCCAGCTGAGATGCA 36113

OY      1276  GTGGGTGTGTCCTCGGCTCAGTGCACCTCTGCTCCAGGTTCGAAGCATTTCTCTG-- 1333
DB      36114  GTGG--GTGATCTCAGCTCAGGCAACCTCTGCTCCAGGTTCGAAGTATCTCTGTGC 36172

OY      1334  TCAGCTTCCCGAATGCTGGAGTATACAGGCGGATGCACACCATGCTTAATTTTGTATT 1393
DB      36173  TCAGCTTCCCAAGTACGTGGAGCTACAGGACAGTGCACACACCGGCTTAATTTGTATT 36232

OY      1394  TTTAGTAGAGACAGAGTTTGCCTATGTTGACCAAGGCTTGCTTGAATCTCTGACTTCAAG 1453
DB      36233  TTTAGTAGAGACAGGTTTTCGCTATGTTGCGCATGTTGGCCAGGCTGTCTTGAATCTCTGACCTCAGG 36292

OY      1454  TGATTCACCCACCTCAGCTCTCCCAAAGCACTGGAGATTACAGGCAATGAGCCACCGTGC 1513
DB      36293  TGATTCACCCACCTCAGCTCTCCCAAAGTGTGGAGATTACAGGCAATGAGCCACCTGCTCCG 36352

OY      1514  GCGTGTTTTTCAGATCCGTGATTTTGTGTTCTGAAAGCTTCATTTGATTTCTT 1567
DB      36353  GCGCGAATTTTCTTTATTTCTTATACATTTTACTTCCGTTGTTTCTTTT 36406

RESULT 6
US-09-949-016-12298
; Sequence 12298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

```

; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CLO01307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12298  
 ; LENGTH: 115963  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(115963)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-12298  
  
 Query Match 10.4%; Score 208; DB 3; Length 115963;  
 Best Local Similarity 79.2%; Pred.No.3,6e-45;  
 Matches 285; Conservative 0; Mismatches 70; Indels 5; Gaps 3;  
  
 Oy 1187 AGAGAAATTTGATTTATTTTTCACATCTTTTATTTTATTTTATTTTAAACAGAG 1246  
 |||||  
 Db 2424 AGAGACACTTACATTTATTTATTTACATTTATTTATTTATTTTATTTTAAACGAA 2483  
 1247 TCTCACTTTGTACCCAGGCTGAGTACAGTGGCTGTGATCTGGCTCACTGCAACTCT 1306  
 2484 TCTCACTGTGTGCCAGGCTGAGTACAGTGG- TGTATCTGGCTCACTGCAACTCTT 2542  
 Oy 1307 GCTTCCCAAGGTTCAAGCATTTCTCTG--TGACTTCCCAATAGCTGGAGTTACAGCG 1364  
 2543 GCTTCCCAAGGTTCAAGCATTTCTCTCTGAGCTCTGAAATACCTGGAGTTACAGGCA 2602  
 Oy 1365 CATGC--ACCAACCAAGCTATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1422  
 2603 CATCCACACACGCTGTGCTATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2662  
 Oy 1423 ACCAGGCTTGGCTTGAATCTCTGACTTCAAGTGCATCCACCACTCAGCTCCCAAGCA 1482  
 2663 GCCAGGCTGTGCTCAAGTCTCTGATGTCAAGTGCATCCCACTCAGCTCCCAAGTG 2722  
 Oy 1483 CTGGAGTTTACAGGATGAGCCAGCGTGGCCAGCTGTTTCTCAGATCCTGTATTTGTTT 1542  
 2723 CTGGAGTTTACAGGATGAGCCAGCTGTGCTGGCCTTCAATTAATTTATGATTTGAGTT 2782  
  
 RESULT 7  
 US-09-949-016-15316  
 ; Sequence 15316, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CLO01307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 15316





```

; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11929
; LENGTH: 92227
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(92227)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11929

Query Match
Best local Similarity 10.3%; Score 206.4; DB 3; Length 92227;
Matches 302; Conservative 0; Mismatches 101; Indels 5; Gaps 3;

QY 1198 TGATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAAGAGTCTCACTTGT 1257
DB 15018 TTATTGATCTTTGGATATATCTTTTATTTTATTTTGAAGAGTCTCACTCTGT 15077
QY 1258 CACCCAGGCTGAGTACAGTGTGCTCTGCTCACTGCAACCTCTGCTCCAGGT 1317
DB 15078 CACCCAGGCTGAGTACAGTGTGCTCTGCTCACTGCAACCTCTGCTCCAGGT 15136
QY 1318 TCAGCGATCTCTCCG--TCAGCTTCCGGAATAGCTGGATTAAGGC--GCATGACCA 1373
DB 15137 TCAGCGATCTCTCCGCTCCGCTCCGATGAGTGGATTAAGCAAGTACCACTAC 15196
QY 1374 CCATGCTTAATTTTGTATTTTATTTTATTTAGTAGAGAGAGTTTGGCAGTTGACCAAGCTTGC 1433
DB 15197 GCCTGGCTAATTTTGTATTTTATTTTATTTAGTAGAGAGAGGTTTCACTGTTGACCAAGCTGT 15256
QY 1434 CTTGAATCTCTGACTTCAAGTATCCACCCACTAGCTTCCAAAGCACTGGATTACA 1493
DB 15257 CTCGAATCTCTGACTTCAAGTATCCACCCACTAGCTTCCAAAGCACTGGATTACA 15316
QY 1494 GGCATGAGCCACCGGCCAGCTGTTTCTCAGATCCCTGATTTGTTTCTGAAGCTTC 1553
DB 15317 GGCATGAGCCACCGGCCAGCTGTTTCTCAGATCCCTGATTTGTTTCTGAAGCTTC 15376
QY 1554 ATTTCTATCTCTTATTTCAATTTTGAAGTATACCTAAGTAAAGTT 1601
DB 15377 ATTAATAATTTCTTCTTACTTTTGTCTCTACTCTGATTAATAAGCTT 15424

RESULT 13
US-09-949-016-15421
; Sequence 15421, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15421
; LENGTH: 92232
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: (1)...(92232)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15421

Query Match
Best local Similarity 10.3%; Score 206.4; DB 3; Length 92232;
Matches 302; Conservative 0; Mismatches 101; Indels 5; Gaps 3;

QY 1198 TGATTATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAAGAGTCTCACTTGT 1257
DB 15018 TTATTGATCTTTGGATATATCTTTTATTTTATTTTGAAGAGTCTCACTCTGT 15077
QY 1258 CACCCAGGCTGAGTACAGTGTGCTCTGCTCACTGCAACCTCTGCTCCAGGT 1317
DB 15078 CACCCAGGCTGAGTACAGTGTGCTCTGCTCACTGCAACCTCTGCTCCAGGT 15136
QY 1318 TCAGCGATCTCTCCG--TCAGCTTCCGGAATAGCTGGATTAAGGC--GCATGACCA 1373
DB 15137 TCAGCGATCTCTCCGCTCCGCTCCGATGAGTGGATTAAGCAAGTACCACTAC 15196
QY 1374 CCATGCTTAATTTTGTATTTTATTTTATTTTATTTAGTAGAGAGTTTGGCAGTTGACCAAGCTTGC 1433
DB 15197 GCCTGGCTAATTTTGTATTTTATTTTATTTAGTAGAGAGGTTTCACTGTTGACCAAGCTGT 15256
QY 1434 CTTGAATCTCTGACTTCAAGTATCCACCCACTAGCTTCCAAAGCACTGGATTACA 1493
DB 15257 CTCGAATCTCTGACTTCAAGTATCCACCCACTAGCTTCCAAAGCACTGGATTACA 15316
QY 1494 GGCATGAGCCACCGGCCAGCTGTTTCTCAGATCCCTGATTTGTTTCTGAAGCTTC 1553
DB 15317 GGCATGAGCCACCGGCCAGCTGTTTCTCAGATCCCTGATTTGTTTCTGAAGCTTC 15376
QY 1554 ATTTCTATCTCTTATTTCAATTTTGAAGTATACCTAAGTAAAGTT 1601
DB 15377 ATTAATAATTTCTTCTTACTTTTGTCTCTACTCTGATTAATAAGCTT 15424

RESULT 14
US-09-949-016-16476/c
; Sequence 16476, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16476
; LENGTH: 141560
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(141560)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16476

Query Match
Best local Similarity 10.3%; Score 206.4; DB 3; Length 141560;
Matches 293; Conservative 0; Mismatches 86; Indels 5; Gaps 3;

QY 1195 ATTTGATTAATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAAGAGTCTCACTT 1254
DB 66406 AATGGAATCGTTTCTTCTTCTTTTATTTTATTTTATTTTGAAGAGTCTCACTC 66347
```

OY	1255	TGTACCCAGAGCTGAGTAACAGGAGCGTGGTCTCGAGCTCAGTCGAACCTCGCCGCCA	1314
Db	66346	TGTACCCAGAGCTGAGTAACAGGAGCGTGGTCTCGAGCTCAGTCGAACCTCGCCGCCG	66288
OY	1315	GGTTCAGCGCATTTCTCCTG--TCAGCTTCCCGAATAGCTGGATTAAG--GCGCATGA	1370
Db	66287	GGTTCAAATGATTTCTCTGCGCTAGCGCTCCCAAGTAGCTGGATTTACAAGTGTGCACAC	66228
OY	1371	CCACCATGCTAATTTTGTATTTTATGTAGAGACAGAGTTTGCAGATGTGACAGGCT	1430
Db	66227	CATGCGTGCTAATTTTGTATTTTATGTAGAGACAGGGGTTCAACATGTGTGCGAGGCT	66168
OY	1431	TGCTTTGAATCTCTGATCTCAGGTGATCCACCACCTCAGCTCCCAAAAGCACTGGAGTT	1490
Db	66167	GGTCTTGAATCTCTGATCTCAGGTGATCCACCACCCTCTGGGCTCTTAAAGTGTGGGATT	66108
OY	1491	ACAGGCAATGAGCCACCGTGGCCAGGCGTTTCTCAGATCTGTATTTGTTCTGAAGCC	1550
Db	66107	ACAGGCGGTGAGCCACCGTGGCCAGGCGGATTTCAAAAATATATCTGTCAATTCAAAATA	66048
OY	1551	TTCAATTTCTAATCTCTTATTCAT	1574
Db	66047	GTACTTAGTGATCATTTTCTTT	66024

Db CTCAACCTCCGACCTCAGAGTATGTGCTCGCTCGGCTCGGCTCCCAAGTCTGAGATTAACA 357  
 QY GGCATGAGCCACCGCGGCCAGCGTGTTTTTCAGATCCGTAATTTGTTCTGAAGCTTC 155  
 Db GGCATGAGCCACCTGCGCCCGCGCTAAATCTTAGTCTTTTAGTAACTTAATTAACAATTT 417  
 QY 1554 ATTCTATCTTCTTAATTATTTGGAAGTAGACCTTAAGTAGGTT 1601  
 Db 418 ATAAATAATTTCTTCTAGATTGTTGCTCCACTCTGAGTAATAAGGTT 465

```

RESULT 16
US-09-949-016-13208/c
; Sequence 13208, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13208
; LENGTH: 97989
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (1) ... (97989)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13208

```



GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 12807  
LENGTH: 23015  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-12807

Query Match 10.3%; Score 205.6; DB 3; Length 23015;  
Best Local Similarity 78.8%; Pred. No. 7.4e-45;  
Matches 271; Conservative 0; Mismatches 69; Indels 4; Gaps 2;

QY 1202 TATGTTTTCAGATCTTTTATTTTATTTTGAAGAGAGCTCACTTGTACCC 1261  
DB 3249 TAAAGGATATCATAGATTTTCTTTTCTGAGACAGAGCTCACTTGTACCC 3190  
QY 1262 CAGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTCAA 1321  
DB 3189 CAGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTCAA 3130  
QY 1322 GCGATTCCTCG--TCAGCTTCCGAAATAGCTGGGATTAACAGGCGCATGC--ACCAACAT 1377  
DB 3129 GTGATTCCTCGCTCAGGCTCCGAGTACGAGGATTAACAGGCGCATGC--ACCAACAT 3070  
QY 1378 GCGTATTTTGTATTTTATTTTATTTTATTTTGAAGAGAGCTCACTTGTACCC 1437  
DB 3069 GCGTATTTTGTATTTTATTTTATTTTATTTTGAAGAGAGCTCACTTGTACCC 3010  
QY 1438 AACTCTGACCTCAGGCAATCCGCGCTCAGGCTCCCAAGTGTGGATTAACAGGCG 1497  
DB 3009 AACTCTGACCTCAGGCAATCCGCGCTCAGGCTCCCAAGTGTGGATTAACAGGCG 2950  
QY 1498 TGAGCCACCGTCCAGGCTGTGTTCTCAGATCCTGTATTTGT 1541  
DB 2949 TGAGCCACCGTCCAGGCTGTGTTCTCAGATCCTGTATTTTGT 2906

RESULT 18  
US-09-949-016-17529/c  
Sequence 17529, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 17529  
LENGTH: 23024  
TYPE: DNA  
ORGANISM: Human

US-09-949-016-17529  
Query Match 10.3%; Score 205.6; DB 3; Length 23024;  
Best Local Similarity 78.8%; Pred. No. 7.4e-45;  
Matches 271; Conservative 0; Mismatches 69; Indels 4; Gaps 2;

QY 1202 TATGTTTTCAGATCTTTTATTTTATTTTGAAGAGAGCTCACTTGTACCC 1261  
DB 3249 TAAAGGATATCATAGATTTTCTTTTCTGAGACAGAGCTCACTTGTACCC 3190  
QY 1262 CAGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTCAA 1321  
DB 3189 CAGGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCAGGTTCAA 3130  
QY 1322 GCGATTCCTCG--TCAGCTTCCGAAATAGCTGGGATTAACAGGCGCATGC--ACCAACAT 1377  
DB 3129 GTGATTCCTCGCTCAGGCTCCGAGTACGAGGATTAACAGGCGCATGC--ACCAACAT 3070  
QY 1378 GCGTATTTTGTATTTTATTTTATTTTATTTTGAAGAGAGCTCACTTGTACCC 1437  
DB 3069 GCGTATTTTGTATTTTATTTTATTTTATTTTGAAGAGAGCTCACTTGTACCC 3010  
QY 1438 AACTCTGACCTCAGGCAATCCGCGCTCAGGCTCCCAAGTGTGGATTAACAGGCG 1497  
DB 3009 AACTCTGACCTCAGGCAATCCGCGCTCAGGCTCCCAAGTGTGGATTAACAGGCG 2950  
QY 1498 TGAGCCACCGTCCAGGCTGTGTTCTCAGATCCTGTATTTGT 1541  
DB 2949 TGAGCCACCGTCCAGGCTGTGTTCTCAGATCCTGTATTTTGT 2906

RESULT 19  
US-09-949-016-15790  
Sequence 15790, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 15790  
LENGTH: 107827  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15790

Query Match 10.2%; Score 204.4; DB 3; Length 107827;  
Best Local Similarity 81.7%; Pred. No. 3.3e-44;  
Matches 273; Conservative 0; Mismatches 56; Indels 5; Gaps 3;

QY 1195 ATTGATTAATGTTTCTCAGATCTTTTATTTTATTTTGAAGAGAGCTCACTT 1254  
DB 74266 ATTGATTAATCTTTCTTTTGTGTTCTGTTTTTTTTTGAAGAGAGCTCACTT 74325  
QY 1255 TGTCAACCAAGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCCA 1314  
DB 74326 TGTCAACCAAGCTGAGTACAGTGGCTGTGCTCGGCTCACTGCAACCTTGCTCCCA 74384  
QY 1315 GGTCAAGGATTCCTCG--TCAGCTTCCGAAATAGCTGGGATTAACAGGCGCATGC--A 1370  
DB 74385 GGTCAAGGATTCATCTCCTCAGCTCTCCAGATGCTGGGATTAACATGATGCGC 74444  
QY 1371 CCACATGCTAATTTTGTATTTTATTTTATTTTGAAGAGAGAGTTTGCCATGTGAACAGGCT 1430

```
Db 74445 CAAGCCACCTATTTTGTATTTTGTAGAGAGAGAGGGGTTTCCATGTTGGTCAAGCT 74504
Qy 1431 TGCCTGAAGTCTCTGACTTCAAGTATCCACCACCTCAAGCTCCCAAGCACTGGAGTT 1490
Db 74505 GGTCTCAAACTCTGACTTCAAGTATCCACCACCTCAAGCTCCCAAGTGTGGAGTT 74564
Qy 1491 ACAGCATGAGCAGCGTCCAGCTGTTTCT 1524
Db 74565 ACAGGTGAGCAGCTGCGCCAGCTCTTTT 74598

RESULT 20
US-09-949-016-13747
; Sequence 13747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13747
; LENGTH: 264665
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13747

Query Match 10.2%; Score 204.4; DB 3; Length 264665;
Best Local Similarity 70.6%; Pred. No. 5.1e-44;
Matches 315; Conservative 0; Mismatches 126; Indels 5; Gaps 3;

Qy 1139 TTAGCCATGTGAGCAAAAGAGAGGCAATTCAGTGTGAATGTTGAGAGAGATTT 1198
Db 23063 TTAGCCATGTGAGCAAAAGAGAGGCAATTCAGTGTTCCTTTGTTGAGAGAGATTT 230122
Qy 1199 GATTAGTTTCTCAGATCTTTTATTTTATTTTGAAGAGAGTCTCAGTTTTC 1258
Db 230123 GATTAGTACATATTTGTTGTTTGTGTTTGTGTTTGAAGAGAGTCTCAGTCTGC 230182
Qy 1259 ACCGAGGTGAGTACAGGTGTGTGTGCTGCTCAGTCAAGCTCTGCTCCAGGTT 1318
Db 230183 ACCGAGTCTAGAGTGTGTGTGAGTCTGCTCAGTCAAGCTCTGCTCCAGGTT 230241
Qy 1319 CAAGCATTTCTCTG--TCAGCTCCGAGATAGTGGATTAAGGCGATGACACACA 1376
Db 230242 CAAGTGAATTTCTGCTCAGCTCTGCTGAGTGGAGTCAAGGCGGCTGACACACA 230301
Qy 1377 --TGCCATATTTTGTATTTTATTTAGAGAGAGAGTGTGCGATGTTGACAGGCTTGC 1434
Db 230302 CCGGCTAATTTTGTATTTTGTATTTGTAAGAGAGAGATTTTGCATGTTGGCCAGGCTATTC 230361
Qy 1435 TTGAATCTCTGACTTCAAGTATCCACCACCTCAAGCTCCCAAGCACTGGAGTTACG 1494
Db 230362 GTGAATCTCTGACTTCAAGTATCCACCACCTTGGCTCCCAAGTGTGGAGTTACG 230421
Qy 1495 GCATGAGCAGCGTCCAGCGCTGTTTCTCAGATCTGATTTGTTTCTGAAGCTTCA 1554
Db 230422 GTGTAGGACCGGTGCGCGAGATTAAGAGATTTTTCAGAGCTCTGTGTTC 230481
Qy 1555 TTTTATCTTTTATTTATTTTGA 1580
Db 230482 ATTTTCTCTTCTGTTTCTTGA 230507
```

```
RESULT 21
US-09-949-016-16874/c
; Sequence 16874, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 16874
; LENGTH: 31867
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16874

Query Match 10.2%; Score 203.8; DB 3; Length 31867;
Best Local Similarity 83.5%; Pred. No. 2.7e-44;
Matches 268; Conservative 0; Mismatches 47; Indels 6; Gaps 3;

Qy 1208 TTCAGATCTTTTATTTTATTTTATTTTGAAGAGAGTCTCACTTTGACCCAGCT 1267
Db 31472 TTCAGATTTTGTGCTTTTGTGTTTGTGTTGAAGAGTCTGCTGTCAACCAAGCT 31413
Qy 1268 GGAGTACAGTGGCTGTGCTGCTCACTGCACTGCTCTGCTCCAGGTTCAAGAT 1327
Db 31412 GGAGTACAGTGGCTGTGCTGCTCACTGCACTGCTCTGCTCCAGGTTCAAGAT 31354
Qy 1328 CTCCTG--TCAGCTTCCGATATGCTGGATTTACAGGCGCATG--CACACCATGCTTA 1382
Db 31353 CTCCTGCTCAGCTCCCGAGTGTGGATTTACAGGACCCGACACCGCCGCTTA 31294
Qy 1383 ATTTTGTATTTTGTGAGAGAGAGTGTGCGATTTGACAGGCTTCCCTTAACATC 1442
Db 31293 TTTTGTATTTTGTGAGAGAGGAGGTTTCAAGTGTGGCAGGCTGTGGAATC 31234
Qy 1443 CTGACTTCAAGTATCCACCACCTCAAGCTTCCCAAGCACTGGATTTACAGGATGAGC 1502
Db 31233 CTGACTTCAAGTATCCGCCACCTCAAGCTTCCCAAGTGTGGATTTACAGGATGAGC 31174
Qy 1503 CACGTTGCCAGCTGTTTC 1523
Db 31173 CACTGTGCCCGCAATTCTC 31153

RESULT 22
US-09-949-016-12864/c
; Sequence 12864, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12864  
LENGTH: 31875  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-12864

Query Match 10.2%; Score 203.8; DB 3; Length 31875;  
Best Local Similarity 83.5%; Pred. No. 2.7e-44;  
Matches 268; Conservative 0; Mismatches 47; Indels 6; Gaps 3;

QY 1208 TTCTCAGATCTTTTATTTTATTTTGAAGAGAGCTCCTCTGACCCAGGCT 1267  
DB 31472 TTCTCAGATTTTGTGCTTTTGTGTTGAAGAGAGCTCCTCTGACCCAGGCT 31413  
QY 1268 GGAGAGAGTGGCTGTGCTCGCTCACTGCACTGCTGCTCCAGGTTCAAGCGATT 1327  
DB 31412 GGAGAGCAATGGC-GTAGTCTGGCTCACTGCACTGCTGCTCCAGGTTCAAGCGATT 31354  
QY 1328 CTCCTG--TCAGCTTCCGAAATAGCTGGATTACAGGCGCATG--CAACGACATGCTTA 1382  
DB 31353 CTCCTGCTCAGCCTCCGAGATAGCTGGATTACAGGCAACCGGCAACGCGCGCTAA 31294  
QY 1383 ATTTTGTATTTTGTAGAGACAGACTTCCCATGTTGACAGGCTTGCCTTGAATCT 1442  
DB 31293 TTTTGTATTTTGTAGAGAGGGGTTTACAGTGTGGCAGGCTGTGTGAACCTC 31234  
QY 1443 CTGACTTCAGGTATCCACCCACTCAGCCTCCCAAGCACTGGATTACAGGATGAGC 1502  
DB 31233 CTGACTTCAGGTATCCGCCCCCTCAGCCTCCCAAGGTCTGGATTACAGGATGAGC 31174  
QY 1503 CACCGTGGCCAGCCTGTTTC 1523  
DB 31173 CACTGTGCCCGGCCAATTCTC 31153

RESULT 23  
US-09-949-016-17593  
Sequence 17593, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17593  
LENGTH: 64024  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-17593

Query Match 10.2%; Score 203.6; DB 3; Length 64024;  
Best Local Similarity 78.2%; Pred. No. 4.2e-44;  
Matches 283; Conservative 0; Mismatches 74; Indels 5; Gaps 3;

QY 1191 AAGTATTTGATTATGTTCTCAGATCTTTTATTTTATTTTGTGAAACAGAGCTTC 1250  
DB 13671 AATATTTTGTAGATTTTATTTTATTTTATTTTGTGATACAGAGCTTC 13730  
QY 1251 ACTTTGTACCCAGGTGAGAGTGTGCTGTGCTGCTCATGCAACTCTGCGCT 1310  
DB 13731 ACTCTGTCCCGAGGCTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13789

QY 1311 CCCAGTTCAAGCATTTCTCTG--TCAGCTTCCGAAATAGCTGGATTACAGGCGCATG 1368  
DB 13790 CCTGTTCAGTATTTCTGTCTCAGCTCTCTGTAGTACTGAGTACAGGCGCAG 13849  
QY 1369 CACCACCAATG-CCTAATTTTGTATTTTGTAGAGACAGAGTTTCCGCAATGTGACCA 1426  
DB 13850 CAACACGCCAGCTAAGTTTGTATTTTGTAGAGATGGGGTTTACCAATGTTGGCA 13909  
QY 1427 GCTTTCCTTGAATCTCTGACTGATGATGATCCACCACTGAGCTCCCAAGCACTGG 1486  
DB 13910 GCTGTCTGCAATCTGCACTGAGTATCCACCACTGAGCTCCCAAGGCTGG 13969  
QY 1487 GATTACAGGATGAGCAACCGTCCGACCTGTTTCTCAATCTGATTTGTTTCTGA 1546  
DB 13970 GATTACAGGATGATGATCAATGCGCGCGCTGATTAATTTTAAATTTTGTCA 14029  
QY 1547 AG 1548  
DB 14030 AG 14031

RESULT 24  
US-09-949-016-12758  
Sequence 12758, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12758  
LENGTH: 88490  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-12758

Query Match 10.2%; Score 203.4; DB 3; Length 88490;  
Best Local Similarity 78.7%; Pred. No. 5.6e-44;  
Matches 281; Conservative 0; Mismatches 71; Indels 5; Gaps 3;

QY 1218 TTTTATTTTATTTTATTTTGTGAAACAGAGTCTCACTTGTCTACCCAGGCTGAGTACAGT 1277  
DB 50407 TTCTATATTTTATTTTATTTTGTAGATGAGTCTGTGTGTACCCAGGCTGAGTACAGT 50466  
QY 1278 GCTGTGTCTGCTCACTGCAACTCTGCTCCAGGTTTAAAGCATTTCTCTG--TC 1335  
DB 50467 GG-TGCAATCTTGGCTCACTGCAACTCTGCTCCAGGTTTAAAGCATTTCTCTGCGC 50525  
QY 1336 AGCTTCCGAAATAGCTGGATTACAGGGCGATGACCAACATG--CTAATTTTGTATT 1393  
DB 50526 AGCTTCCGAAATAGCTGGATTACAGGGCTGCTGCAACACAGCGGCTTAAGTTTGTATT 50585  
QY 1394 TTTTATGAGACAGAGTTTCCGATGTTGACCAAGGCTTGTGAACTTCTGACTTCAGG 1453  
DB 50586 TTTTGTGAGAGAGAGTTTGTGCAATGTTCCAGGGTGTCTTGAATCTCTGACCTCAGG 50645  
QY 1454 TGATCCACCACCTGAGCTCCCAAGCACTGGATTACAGGATGAGGACCGGCGCA 1513  
DB 50646 TGATCCACCACCTGAGCTCCCAAGGCTGGATTACAGGATGAGGACCGGCGCA 50705  
QY 1514 GCTGTCTTCTAGATCTGTATTTTGTGAAAGCTTCAATTTCTATCTTAT 1570  
DB 50706 GCTTATATCTTATTTATTAACATGTTGATTAATAATTTATCTACAAATTTTAT 50762

```

RESULT 25
US-09-949-016-14222
; Sequence 14222, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14222
; LENGTH: 88736
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14222

Query Match 10.2%; Score 203.4; DB 3; Length 88736;
Best Local Similarity 78.7%; Pred. No.5,6e-44;
Matches 281; Conservative 0; Mismatches 71; Indels 5; Gaps 3;

QY 1218 TTTTATTTTATTTTATTTTATTTTGAACAGAGCTCTCACTTTGTCAACCAGGCTGAGTACAGT 1277
DB 50653 TTCAATAATTTTTTTTTTTTTTTTGGAGATGAGACTCTTGCTTCGTACCCAGGCTGAGTACAGT 50712

QY 1278 GGGCTGTGGCTCGGCTGCTACGTCAACCTCTGCTCCCGAGTTCAAGGATTTCTCTCG--TC 1335
DB 50713 GG-TGCGATCTTGGCTCACTGCAACTCTGCTCCCGAGTTCAAGGATTTCTCTCGCGC 50771

QY 1336 AGCTTCCCGAATAGCTGTGAGATTACAGCGCCATGACACCACTATG--CTAATTTTGTATT 1393
DB 50772 AGCTTCCCGAATAGCTGTGAGATTACAGCGCCATGACACCACTATGAGTTTGTATT 50831

QY 1394 TTTTATGAGAGACAGAGTTTGGCCATGTGTACCAAGGCTTGGCTTGAACCTCTGACTTCAAG 1453
DB 50832 TTTTGTGAGAGACGAGATTGTTCATCATGTTCGCCAGGGTGTGCTTGAACCTCTGACTTCAAG 50891

QY 1454 TGATGCACCCCACTCAGCTCCCAAGACGACTGGAGATTACAGGATAGGACACCGTSCCA 1513
DB 50892 TGATTCGCCCACTCGGCTCCCAAGAGTCTGGAGATTACAGGATAGGACACCGTSCCG 50951

QY 1514 GCGTGTTTTTCAGATCCTGTGATTTGTTCGTAAGCCTTCATTCTCTTATT 1570
DB 50952 GCGTGTATTAATTATTATTATAACATGTTGATTAATAAATTATCTACATATTATTATT 51008

RESULT 26
US-09-949-016-11749/c
; Sequence 11749, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

```

```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11749
; LENGTH: 144596
; TYPE: DNA
; ORGANISM: Human
FEATURES:
NAME/KEY: misc_feature
LOCATION: (1)...(144596)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11749

Query Match      10.1%; Score 202.8; DB 3; Length 144596;
Beet Local Similarity 78.5%; Pred. No. 1e-43; Mismatches 72; Indels 5; Gaps 3;
Matches 281; Conservative 0;

Db          97169 TGGCAATATTACTTTTAAATTTTATTTTATTTTATTTTATTTGAGA 97110
Oy         1182 TGTTAGAGGAAGTATTGATTATGTCTCAGATCTTTATTTTATTTTGAAA 1241
           |||||
Db          97169 TCGCAATAATTACTTTTAAATTTTATTTTATTTTATTTTATTTTATTTGAGA 97110
Oy         1242 CAGAGTCATCTTGTCTCACCCAGGCTGAGTAGACAGTGGCTGTGCTGCAGTCGAA 1301
           |||||
Db          97109 TGAATCTCATCTGTGCCGCCAGGCTGAGTGCAGTGGC-GGAGTCTTGAGCTCAGTCAA 97051
Oy         1302 CCTTCGCTCCCGAGGTTCAAGGATTTCTCG--TCAGTTCGCCAATAGTGGGATTAC 1359
           |||||
Db          97050 CCTTCGCTCCCGAGGTTCAACAATTCATCTCAGCTCCGAGTGTGGGATTAC 96991
Oy         1360 AGGCGCATGACACCACCATGCG--CTAATTTTTGTATTTAGTAGAGACAGAGTTGCGCA 1417
           |||||
Db          96990 AGGCGCATGACACACACCCAGCTAATTTTGTATTTTATTTAGTAGACAGGAGTTTACCA 96931
Oy         1418 TGTTGACAGGAGCTTGCTTGAATCTCTGACTTCAGGTATTCACCCAGCTCAGCTCCCA 1477
           |||||
Db          96930 TGTGTGTAAGGTGTGTCTTGAATCTGACTGACTCAAGTATTCACCACTCGGCTCCCA 96871
Oy         1478 AAGCACTGGGATTACAGCATGAGCAACCGTCCAGCTGTTTCTCAGATCTGTGA 1535
           |||||
Db          96870 AAGTGCTGGGATTACAGGCGTGAGCCACCATGACTGGCTTAATTTTACTTTTGTGA 96813
Oy         1478 AAGTGCTGGGATTACAGGCGTGAGCCACCATGACTGGCTTAATTTTACTTTTGTGA 96813

RESULT 27
US-09-949-016-13035/c
; Sequence 13035, Application US/09949016
; Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949, 016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13035
LENGTH: 144596
TYPE: DNA
ORGANISM: Human
FEATURES:
NAME/KEY: misc_feature
LOCATION: (1)...(144596)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13035

Query Match      10.1%; Score 202.8; DB 3; Length 144596;
Beet Local Similarity 78.5%; Pred. No. 1e-43; Mismatches 72; Indels 5; Gaps 3;
Matches 281; Conservative 0; Mismatches 72; Indels 5; Gaps 3;

```

QY 1182 TGTTCAGAGAGATATTTGATTTCTCAGATCTTTTATTTTATTTTGTGAAA 1241  
DB 97169 TGGCAAAATATTTACTTTTAAATTTTATTTTATTTTATTTTATTTTGTGAGA 97110  
QY 1242 CAGAGTCTCAGTTTGTCTCACCAGGCTGAGTACAGTGGCTGTCTGCGCTCATCGCAA 1301  
DB 97109 TGGATCTCAGTGTCTGCGCCAGGCTGAGTACAGTGGC-GCAAGCTTGTGCTCAGTGGCAA 97051  
QY 1302 CCTGCGCTCCAGGTTCAAGGATTTCTCTG--TCAGCTTCCGAATAGTGGGATTAC 1359  
DB 97050 CCTGCGCTCCAGGTTCAAAATTTCTATGCTCAGCTCCGAGTACGCTGGGATTAC 96991  
QY 1360 AGGCGCATGACACACCATTGC--CTAATTTTGTATTTTATAGAGACAGATTTGCGCCA 1417  
DB 96990 AGGCGCATGACACACCATTGTTTGTATTTTATAGTACAGGCGGTTTCACCA 96931  
QY 1418 TGTTCACAGGCTTGTGCTTGAATCTCTGACTTCAGTATCCACCACTCAGCTCCCA 1477  
DB 96930 TGTTCACAGGCTTGTGCTTGAATCTCTGACTTCAGTATCCACCACTCAGCTCCCA 96871  
QY 1478 AAGCATGGGATTATACAGCATGAGCCAGCGTCCAGCTGTTTCTCAGATCTGTGA 1535  
DB 96870 AAGCATGGGATTATACAGCGCTGAGCCACCATGATGCGCTAAATTTTATCTTTGTGA 96813

## RESULT 28

US-09-949-016-15681/c  
; Sequence 15681, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/231,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 15681  
; LENGTH: 80269  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15681

Query Match 10.1%; Score 202.6; DB 3; Length 80269;  
Best Local Similarity 75.6%; Pred. No. 8.8e-44;  
Matches 291; Conservative 0; Mismatches 89; Indels 5; Gaps 3;

QY 1217 CTTTATTTTATTTTATTTTATTTTGAACAGATCTCATTTGTACCCAGGCTGAGTACAG 1276  
DB 13610 CATTTTATTTTATTTTATTTTATTTTGAAGAGATCTCTGCTGTCTGTACCCAGGCTGAGTACAG 13551  
QY 1277 TGGCTGTGCTCTGCTCACTGCACTTGTCTGCTTCCAGGTTCAAGCATTTCTCTG--T 1334  
DB 13550 TGGC-ATGACCTGCTCTCTCAACACTTGTCTGCTCTGCTGAGTTCAAGTCTCTGCTCT 13492  
QY 1335 CAGCTTCCGATAGCTGTGGATTAACAGGCGCATGC--ACCACATGCTCTAATTTTGTAT 1392  
DB 13491 CAGCTTCCGATAGCTGTGGATTAACAGCTGCGCCACACAGCGCTGATTTTGTAT 13432  
QY 1393 TTTTATAGAGACAGATTTGCGCATGTTGACAGGCTTGCCTGAATCTCTGACTTCAG 1452  
DB 13431 TTTTATAGAGACAGATTTGCGCATGTTGACAGGCTTGTGCTGCTCAATCTCTGACTTCAG 13372  
QY 1453 GTGATCCACCACTGAGCTTCCAAAGCACTGGAATTACAGGCTATGAGCCACCGTGGCC 1512

DB 13371 ATGATCACCCGCTCAGCTCCCAAGAGGCTGGGATTACAGGATGACCAAGCTGGCC 13312  
QY 1513 AGCGTGTTCAGATCTGTATTTGTTCGAAAGCTTCATTTCTATTTATTTCA 1572  
DB 13311 AGCGATGATGACATTTCTACAGTCTCTATGAGATGATATTTTGTAAAAA 13252  
QY 1573 TTTTGAAGTATGACCTTAAGTAA 1597  
DB 13251 AGTGTCTTATGATTAATTTAAGCAA 13227

## RESULT 29

US-09-949-016-15870/c  
; Sequence 15870, Application US/09949016  
; Patent No. 6812339

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: CL001307  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 15870  
; LENGTH: 45418  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15870

Query Match 10.1%; Score 202.4; DB 3; Length 45418;  
Best Local Similarity 77.7%; Pred. No. 7.5e-44;  
Matches 283; Conservative 0; Mismatches 76; Indels 5; Gaps 3;

QY 1192 AGTATTTGATTTATTTTCTCAGATCTTTTATTTTATTTTGAACAGATCTCA 1251  
DB 15216 ATTTCTGTTTTTTTTTGTGTTGTTGTTTTGTTTTTTTTTGTGACAGATCTCG 15157  
QY 1252 CTTTGTCAACCCAGGCTGAGTACAGTGTGCTGTGCTGCTCACTGCACTCTGCTC 1311  
DB 15156 CTCTGTGCGCCAGGCTGAGTACAGTGTGCG--GCAATCTTGGCTCACTGCACTCTGCTC 15098  
QY 1312 CAGGTTCAAGGATTTCTCTG--TCAGCTTCCGAATAGCTGGGATTAACAGGCGCATGC 1369  
DB 15097 CCGAGTTCAAGGATTTCTCTGCTCTGCTCTCCCAAGTATGCTGGGATTAACAGGCGCCAC 15038  
QY 1370 ACCACATGCG--CTAATTTTGTATTTTATAGAGACAGATTTGCGCATGTGACCGAG 1427  
DB 15037 CATCAAGTTCAAGTATTTTGTATTTTATAGAGACAGGATTTGCGCATGTGCTGAG 14978  
QY 1428 GCTTGTCTGAACTCTGCTCACTTCAAGTATCAACCACTCAGCTTCCAAAGCACTGGG 1487  
DB 14977 GCTTGTCTGAACTCTGCTCACTTCAAGTATCAACCACTCAGCTTCCAAAGTGTGAGT 14918  
QY 1488 ATTTACAGGATGAGCAACCGTCCAGGCTGTTTCTCAGATCTGTATTTTGTGAA 1547  
DB 14911 ATTTACAGGATGAGCAACCGTCCAGGCTTGTGTTCTTATATTAATTAATTTATGGA 14858  
QY 1548 GCCT 1551  
DB 14857 CCGT 14854

## RESULT 30

US-09-949-016-16238/c  
; Sequence 16238, Application US/09949016  
; Patent No. 6812339

```
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 16238
/ LENGTH: 59828
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(59828)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16238

Query Match          10.1%; Score 202.4; DB 3; Length 59828;
Best Local Similarity 78.7%; Pred. No. 8.6e-44;
Matches 280; Conservative 0; Mismatches 71; Indels 5; Gaps 3;

QY 1174 GTTGAATGTTTCAGAGATATTTGATATGTTTCTCAGATCTTTTATTTTATTTT 1233
DB 52769 GTGAGATTTGAGAGATTTGATATGTTTCTCAGATCTTTTATTTTATTTT 52710
QY 1234 TTTTGAACAGAGTCTCACTTGTGACCCAGGCTGAGTACAGTGTGCTGTGCT 1293
DB 52709 TTTTGAACAGAGTCTCACTTGTGACCCAGGCTGAGTACAGTGTGCTGTGCT 52651
QY 1294 CACTGCAACTGCTGCTCCAGGTTCAAGGATTCCTCCG--TCAGCTCCGGAATGCT 1351
DB 52650 TACTGCAACTGCTGCTCCGAGTTCAAGGATTCCTCCGCTCCGAGTACGCT 52591
QY 1352 GGGATTACAGGCGCATGC--ACCACCATGCTAATTTTGTATTTTGTATAGACAGAG 1409
DB 52590 GGGATTACAGGCGGTGTGACCAAGCCGCTAATTTTGTATTTTGTATAGAGGCGG 52531
QY 1410 TTTTCCCATGTTGACCAAGGCTTGCCTTGAATCTCTGACTTCAAGTGATCCACCACTTA 1469
DB 52530 TTTTCCCATGTTGACCAAGGCTTGCCTTGAATCTCTGACTTCAAGTGATCCACCACTTA 52471
QY 1470 GCTTCCCAAGGCTGAGTTACAGGATGAGGACCCGCGCCAGGCTGTTTCTC 1525
DB 52470 GCTTCCCAAGGCTGAGTTACAGGATGAGGACCCGCGCCAGGCTTTCATATC 52415

RESULT 31
US-09-949-016-12257
/ Sequence 12257, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15368
/ LENGTH: 129327
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(129327)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15368

Query Match          10.1%; Score 202.4; DB 3; Length 129327;
Best Local Similarity 81.6%; Pred. No. 1.3e-43;
Matches 271; Conservative 0; Mismatches 56; Indels 5; Gaps 3;
```

```
/ SEQ ID NO 12257
/ LENGTH: 129327
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(129327)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12257

Query Match          10.1%; Score 202.4; DB 3; Length 129327;
Best Local Similarity 81.6%; Pred. No. 1.3e-43;
Matches 271; Conservative 0; Mismatches 56; Indels 5; Gaps 3;

QY 1195 ATTGATATGTTTCTCAGATCTTTTATTTTATTTTATTTTGAACAGATCTCACTT 1254
DB 29960 ATCTAGATATTTTACCTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 30019
QY 1255 TGTCAACCCAGGCTGAGTACAGTGTGCTGCTGCTCACTGCAACCTGCTCCCA 1314
DB 30020 TGTCCCTGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 30078
QY 1315 GGTTCAGAGATTTCTCTG--TCAGCTTCCGAAATAGTGGATTTACAGGCGCATGC--A 1370
DB 30079 AGTTCAAGTATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 30138
QY 1371 CCACCATGCTTATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1430
DB 30139 CATGCTGCTTATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 30198
QY 1431 TGCCTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1490
DB 30199 GCTTCAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 30258
QY 1491 ACAGCATGAGCCACCGTCCAGGCTGTTT 1522
DB 30259 ACAGGTGAGGCAACCGTCCAGGCTGTTT 30290

RESULT 32
US-09-949-016-15368
/ Sequence 15368, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15368
/ LENGTH: 129327
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(129327)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15368

Query Match          10.1%; Score 202.4; DB 3; Length 129327;
Best Local Similarity 81.6%; Pred. No. 1.3e-43;
Matches 271; Conservative 0; Mismatches 56; Indels 5; Gaps 3;

QY 1195 ATTGATATGTTTCTCAGATCTTTTATTTTATTTTATTTTATTTTGAACAGATCTCACTT 1254
```





```

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 17521
; LENGTH: 87734
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17521

```

Query Match	10.1%;	Score 201.6;	DB 3;	Length 87734;
Best Local Similarity	81.0%;	Pred. No. 1.7e-43;		
Matches 272;	Conservative	0;	Mismatches 59;	Indels 5;
			Gaps	3;

Qy	1306	TTTTTCAGATCTTTTAAATTTTAAATTTTGTGAACAGAGTCTCACTTTGTACACCAAG	1265
Db	26852	TCCTTCAGTTCTCTTTTTTTGTTTTTTTTTGAACAGAGTCTCCTCTCCGACACCAAG	26793
Qy	1266	CTGAGATGACAGTGGCTGTGCTCTGCGCTCACTGCAACTTGTGCTCTCCAGGTTCAAGCA	1325
Db	26792	CTGAGATGCAATGGC--GTGATCTCGGCTCTCTGCAATCTTCAACTCCAGGTTCAAGCA	26734
Qy	1326	TTTCCTCG--TCAGCTTCCGGAATAGCTGGAGATTACAGGCGCATGACACCACT--GCCT	1381
Db	26733	TTCTCTCGCTCAACCTCCGAGTAGCTGGAGCTAACAGGACGTTGCAACATCCGGCT	26674
Qy	1382	AATTTTGTATTTTAAAGAGACAGATTTCGCATGTTGACCAAGGCTTGCTTGAATCT	1441
Db	26673	AATTTTGTATTTTAAATTAAGACAGGATTTCAACAATTTGTGAAGCTGTCTCAACT	26614
Qy	1442	CCTGACTTCAGGTGATCCACCACTTCAAGCTCCCAAGCACTGGGATTACAGGCAATGAG	1501
Db	26613	CCTGACCTCAGGTGATCCACCTGCTTCAAGCTCCCAAAAGTGTGGGATTACAGGTGTGA	26554
Qy	1502	CCACCGTGCCAGGCTGTTTTCTTCACATCTGTATTT	1537
Db	26553	CCACCATGCAAGGCTCTTCTCAACAAGTTTCTAGT	26518

RESULT 36  
 US-09-949-016-13469/C  
 ; Sequence 13469, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CLO01307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13469  
 ; LENGTH: 46347  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-016-13469

Query Match	10.1%;	Score 201.4;	DB 3;	Length 46347;
Best Local Similarity	85.0%;	Pred. No. 1.4e-43;		
Matches 261;	Conservative	0;	Mismatches 41;	Indels 5; Gaps 3;

QY	1216	TCCTTTATTTTATTTTATTTTATTTTAAACAGAGTCTCACTTGTACACCCAGAGCTGAGATCA	1275
Db	35745	TCCTTCTTTTATTTTATTTTATTTTATTTTGAAGAGATTGTCTTTGTACCCAGCTGAGAGTGA	35686
QY	1276	GTGGCTGTGGTCTCGGCTCACTGCACTCTGCTGCCCTCCAGTTCAAGCATTTCTCGT--	1333
Db	35685	GTGGC-ACGATCTTGGCTACTGCAACTCTGTCTCCAGTTCAATATTTCTCTGCC	35627
QY	1334	TCAGCTTCCGGAATAGCTGGGATTTACAGGCGCATGCAACCAACATG--CTAATTTTTGTGA	1391
Db	35626	TCAGCTTCCGAGTAGCTGGGATTTACAGGCGCTGCACACACCAGCTAATTTTTTGTGA	35567
QY	1392	TTTTTATGTAAGACAGAGTTTGGCCATGTTGACCAAGCTTGGCTTGAATCTCTGACTTCA	1451
Db	35566	TTTTTATGTAAGACAGAGGTTTCAACATGTTTGGCCAAATTGGTCTTGAATCTCTGCACTTCA	35507
QY	1452	GGTATATCAACCAACTCAGCTTCCCAAGACATGGGATTTACAGGCAATGAGCCACCGTGGC	1511
Db	35506	GGTATATCAACTGCTCCTCAGCTTCCCAAGATGTGGGATTTACAGGTGTGAGCACCGCGCC	35447
QY	1512	CAGGCTTG 1518	
Db	35446	CAGGCTTG 35440	

```

RESULT 37
US-09-949-016-15830/c
; Sequence 15830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15830
; LENGTH: 192506
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15830

```

Query Match	10.1%;	Score	201.4;	DB	3	Length	192506;
Best Local Similarity	73.0%;	Pred.	20.2;	8e-43			
Matches	300;	Conservative	0;	Mismatches	106;	Indels	5;
						Gaps	3;
QY	1169	TCAGTGTGAATGTTCA	GAGAGTAATTGA	TATATGTTTTCT	CAGATCTTTTATATTT		1228
Db	76912	TCAAGGTTTATATTTGG	TGTCATTTTCTTTTCTTTTCTGTTTTTTTTTTTTT				76853
QY	1229	ATTTTTTTTAAACAGAG	CTCACTTGTACCCAGGCTGAGATA	CAGTGC	-TGTGGTC		1287
Db	76852	TTTTTTTTTGAAGACAG	AGTCTCACTGTGCACCCAGCTTGA	GTGATGTGCA	GTGCACATC		76793
QY	1288	TCGGTCACTGCACACCT	CTGTCTCCAGGTTCAAGCATCTCTG	-TCACTTCCGA			1345
Db	76792	TCAGCTCACTGTAACCT	CCACCTTCAGGTTCAAGAAATCTCTGCTCA	GCCTCCGA			76733
QY	1346	ATAGCTGGGATTCAGAG	CGCATGC-ACACATGCTGTAATTTTGTATTTT	TAGTAG			1403
Db	76732	GTMGCTGGGATTTAAG	CGCATGTCTGCCATATTTTGTATTTT	TAGTAAG			76673
QY	1404	ACAGAGTTTGGCATGTT	GAACCAAGCTTGCTTGAATCTCTGAC	TTCAAGTATCAACC			1463
Db	76672	ACAGGGTTTACCATATG	TGGCCAGGCTGTGATCTTGAATCTCTG	GCCTCAAGTGATCAACC			76613



1193 GTATTGATTTATGTTTCTCAGATCTTTTATTTTATTTTGAACAGAGTCTCAG 1252  
14250 GATTTTCTTTTGGTTTCTTCTCTTTTATTTTATTTTATTTTGAACAGAGTCTC 14191  
1253 TTTGTCACCAAGGCTGAGTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1312  
14190 TCTGTACCAAGGCTGAGTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 14132  
1313 CAGGTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1370  
14131 CGGGTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14072  
1371 CCACCATGCG--CTAATTTTGTATTTTGTAGAGACAGATTTTGGCAGTGGACAG 1428  
14071 ACCACACCCCACTAATTTTGTATTTTGTAGAGACAGATTTTGGCAGTGGACAG 14012  
1429 CTTCCTTGAATCTCTGACTGAGTATGACCACTGAGCTTCCCAAGCAGTGGAG 1488  
14011 CTGGTCTGAACTCTCGACTGAGTATGACCACTGAGCTTCCCAAGTGTGGGA 13952  
1489 TTACAGGATGAGCCAGCGTCCAGCTGTTT 1522  
13951 TTACAGGCGTGAAGCCAGCGCTGCTT 13918

Query Match 10.1%; Score 201.2; DB 3; Length 99749; Best Local Similarity 81.1%; Pred. No. 2.3e-43; Matches 271; Conservative 0; Mismatches 58; Indels 5; Gaps 3;

1193 GTATTGATTTATGTTTCTCAGATCTTTTATTTTATTTTGAACAGAGTCTCAG 1252  
14250 GATTTTCTTTTGGTTTCTTCTCTTTTATTTTATTTTGAACAGAGTCTCAG 14191  
1253 TTTGTCACCAAGGCTGAGTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1312  
14190 TCTGTACCAAGGCTGAGTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 14132  
1313 CAGGTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1370  
14131 CGGGTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14072  
1371 CCACCATGCG--CTAATTTTGTATTTTGTAGAGACAGATTTTGGCAGTGGACAG 1428  
14071 ACCACACCCCACTAATTTTGTATTTTGTAGAGACAGATTTTGGCAGTGGACAG 14012  
1429 CTTCCTTGAATCTCTGACTGAGTATGACCACTGAGCTTCCCAAGCAGTGGAG 1488

US-09-949-016-16518/c  
Sequence 16518, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16518  
LENGTH: 99749  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-16518

14011 CTGTCTGAACTCTGAGCTTCAAGTATGACCACTGAGCTTCCCAAGTGTGGGA 13952  
1489 TTACAGGATGAGCCAGCGTGGCAGCTGTTT 1522  
13951 TTACAGGCGTGAAGCCAGCGCTGCTT 13918

Query Match 10.1%; Score 201.2; DB 3; Length 162025; Best Local Similarity 66.7%; Pred. No. 2.9e-43; Matches 348; Conservative 0; Mismatches 168; Indels 6; Gaps 4;

1197 TTGATTATGTTTCTCAGATCTTTTATTTTATTTTGAACAGAGTCTCAGT 1256  
56490 TGGTTATGATTTGAGTATTTTGGTTTGTGTTTGTGTTTGAACAGTCTTACTG 56549  
1257 TCACCCAGGCTGAGTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1316  
56550 TCACCCAGGCTGAGTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 56608  
1317 TCACCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1374  
56609 TTTGAGCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 56668  
1375 CA--TGCTTAATTTTGTATTTTGTAGAGACAGATTTTGGCAGTGGACAGCT 1432  
56669 CACCGGCTTAATTTTGTATTTTGTAGAGTGGTTCACATGTTGGCAGGTTG 56728  
1433 CTTGAACTCTGACTTGAAGTATGACCACTGAGCTTCCCAAGCAGTGGATTAC 1492  
56729 TCTTAACCTCTGAGCTGAGTATGACCACTGAGCTTCCCAAGGCTGAGATTAC 56788  
1493 AGGATGAGCCAGCGTGGCAGCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1552  
56789 AGGCTGAGCCAGCGGCTGGCAGGTTTGAAGTATTTATCTGATTTTATTAAGAT 56848  
1553 CATTTCTATCTTATTTATTTTGAAGTATGACCACTGAGTATTTTGAACATCA 1612  
56849 AGTGTATGTTTGTGTTCTTATTTTCTGATCTGTTGGTGA-ATTGTAAGATCCC 56907  
1613 AATATCTTGAAGAAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1672  
56908 CATATCTACACAGAGTCAATTTCAATCTTTTCTGAGCTGTTATTTATTTAT 56967  
1673 CTGATGTTATGTTTCTTCAATTTATTTATTTATTTATTTATTTATTTATTTAT 1714

US-09-834-700-13  
Sequence 13, Application US/09834700  
Patent No. 6958214  
GENERAL INFORMATION:  
APPLICANT: Braun, A.  
TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING THE SAME  
FILE REFERENCE: 24736-2035  
CURRENT APPLICATION NUMBER: US/09/834,700  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: 60/217,251  
PRIOR FILING DATE: 2000-07-10  
PRIOR APPLICATION NUMBER: 60/240,335  
PRIOR FILING DATE: 2000-10-13  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 162025  
TYPE: DNA  
ORGANISM: Homo Sapien  
FEATURE:  
NAME/KEY: allele  
LOCATION: 83587  
OTHER INFORMATION: Nucleotide sequence of chromosome 17 with genomic  
OTHER INFORMATION: sequence of the allelic variant AKAP10-6  
US-09-834-700-13



```

Patent No 6958214
GENERAL INFORMATION:
APPLICANT: Braun, A.
TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: 24736-2035
CURRENT APPLICATION NUMBER: US/09/834,700
CURRENT FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/217,251
PRIOR FILING DATE: 2000-07-10
PRIOR APPLICATION NUMBER: 60/240,335
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 162025
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: allele
LOCATION: 156,217
OTHER INFORMATION: Nucleotide sequence of chromosome 17 containing
OTHER INFORMATION: the genomic sequence of the allelic variant
US-09-834-700-18

```

Query Match	10.1%	Score 201.2	DB 3	Length 162025
Best Local Similarity	66.7%	Pred. No. 2.9e+3		
Matches 348	Conservative 0	Mismatches 168	Indels 6	Gaps 4

QY	1197	TTGATTAATGTTTCTCAGATCTTTTATTTTATTTTATTTTGGAAAAGAGTCTCACTTGG	1256
Db	56490	TGGTTTATTAATTTTGAAGATTTTGGTTTGGTTTGGTTTGGTTTGGAGAAGTCTTACTCTGG	56549
QY	1257	TCACCCAGGCTGGAGTACAGTGGCTGGGTCTGGCTCACTGCAACTCTGACTCCCAAG	1316
Db	56550	TCACCCAGGCTGGAGAGCAAGTGGC-AATAATCAAGTCACTGCAACTCTGTCTCCCAAG	56608
QY	1317	TTCAAGGGAATTTCTCTGG--TCAGCTTCCCAATAGCTGGGATTTACAGGCGCATGCACCAc	1374
Db	56609	TTTGAAGGATTTCTCTGCTCCTCAGGCTCACTGATAGTGGGATTTACAGGCGCGTGCACCA	56668
QY	1375	CA--TGCTTAATTTTGTATTTTATTTAGTAGAGACAGATTTCCGCATGTTGAACAAGCTTGG	1432
Db	56669	CACCCGCTAATTTTGTATTTTGTATTTTATTTAGTAGAGATGGGGTTTCAACATGTTGGCCAGGGTGG	56728
QY	1433	CTTTGAATCTCTCACTTCAAGATATCCACCACTCACTCAAGCTCCCAAGAACAAGCTGGATTAC	1492
Db	56729	TCTTGAATCTCTCACTCAAGATATCCACTGCTCAAGCTCCCAAGAAGCTGAGATTAC	56788
QY	1493	AGGATATAGGCAACGATGCCAGCTGTTTTCTCAGATCCTGTATTTTGTCTTGAACCTT	1552
Db	56789	AGGAGTAGGCAACCGGCTCTGGCATGGTTTGGAGATTTAATCTGTAGTTTATTAATAAGAT	56848
QY	1553	CATTTCATCTTCTTATTCATTTTGGAAAGTAGTACAACTAAGTAGGTTTTTAACATCA	1612
Db	56849	AGTCTTATAGTTGGTTGTCTTATATTTCTTGGTAACTTGGGTA-ATTTGTAAAGTACC	56907
QY	1613	AATATCTTTGGAAAATTCCTGGTCTTCTTCTTATTTCTTACAAAATATGTTCAGTATAG	1672
Db	56908	CATATCTACACAAGAAGTCAATTTTCAATCTTTTCTTCAGACGTGTTATTTATTTAT	56967
QY	1673	CTGATGTATGTTCTTCAAAATATTCATTTCTTATCTCA	1714
Db	56968	TTTATTTTATTTTATGTTTGAATGAGAGTCTGCTGTGTCA	57009

Search completed: January 22, 2006, 21:30:33  
Job time : 257.224 secs

(c) (5) DPP, (c) (5) ACP

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5419.54 Seconds  
(without alignments)  
17274.699 Million cell updates/sec

Title: US-09-728-552a-3\_COPY\_39000\_41000

Perfect score: 2001

Sequence: 1 tgcaccacgacgttcgctc.....cgaatcagtggaacttca 2001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	417.2	20.8	494	9	AQ598756 HS_5338.B
2	214.4	10.7	744	9	BZ610799 WHADJ52TR
3	212.8	10.6	871	9	BZ603854 WHADJ52TR
4	211.4	10.6	658	7	CN266085 WHADJ52TR
5	208.8	10.4	443	2	BF991881 PM4-GEN001
6	208.2	10.4	435	2	BF991882 PM4-GEN001
7	206.8	10.3	643	1	AM975010 ESTJ387115
8	206.4	10.3	683	1	AM975010 ESTJ387115
9	206.4	10.3	361	1	AM572721 h686b11.x
10	206.4	10.3	866	9	AO889099 HS_5263.A
11	206	10.3	605	9	AQ347764 RCT11-13
12	206	10.3	812	2	BG026977 602294004
13	205.6	10.3	710	5	B0849843 AGENCOURT
14	205.4	10.3	544	1	AU147414 AU147414
15	205.2	10.3	2768	4	CR936687 Homo sapi
16	205	10.2	455	9	AO037381 CIT-HSP-2
17	205	10.2	462	6	CD699526 ESTJ16162
18	205	10.2	693	9	BZ612424 WHADJ52TR
19	205	10.2	704	2	BE787136 BE787136
20	204.4	10.2	555	5	BX484164 DKFZP686J
21	204.4	10.2	829	9	BZ599806 WHADJ62TR
22	204.2	10.2	458	1	AI733856 zol19c03.y

23	203.6	10.2	390	1	AA808875 nt86h02.s
24	203.6	10.2	438	7	CR739570 CR739570
25	203.4	10.2	545	2	BF854170 BF854170
26	203	10.1	526	11	CR961104 CR961104
27	203	10.1	3165	4	CR859576 CR859576
28	202.8	10.1	492	6	CD701193 CD701193
29	202.8	10.1	888	5	BU957691 BU957691
30	202.6	10.1	742	9	AO582332 AO582332
31	202.4	10.1	470	10	AG190706 AG190706
32	202.2	10.1	449	9	AQ769842 AQ769842
33	202.2	10.1	466	9	AQ822778 AQ822778
34	202.2	10.1	552	1	AI479068 AI479068
35	202.2	10.1	552	1	AI712937 AI712937
36	202.2	10.1	930	5	BU517124 BU517124
37	202	10.1	935	2	BF668880 BF668880
38	201.8	10.1	439	1	AI712324 AI712324
39	201.8	10.1	662	10	AG037584 AG037584
40	201.8	10.1	678	1	AV729893 AV729893
41	201.8	10.1	1102	3	BM458632 BM458632
42	201.6	10.1	456	1	AM408767 AM408767
43	201.6	10.1	524	1	AI791156 AI791156
44	201.6	10.1	714	2	BG545785 BG545785
45	201.6	10.1	3181	4	CR936826 CR936826
46	201.2	10.1	711	10	AG010148 AG010148
47	201.2	10.1	730	6	CB308016 CB308016
48	201	10.0	749	6	CG962143 CG962143
49	200.8	10.0	692	10	AG173021 AG173021
50	200.8	10.0	696	5	BU616112 BU616112
51	200.6	10.0	439	1	AI707781 AI707781
52	200.6	10.0	618	6	CD579489 CD579489
53	200.6	10.0	618	6	CD579508 CD579508
54	200.6	10.0	861	6	CD517715 CD517715
55	200.6	10.0	907	6	CD558895 CD558895
56	200.4	10.0	490	1	AI792575 AI792575
57	200.4	10.0	690	10	AG182694 AG182694
58	200.4	10.0	739	5	BU153359 BU153359
59	200.2	10.0	548	1	AM962006 AM962006
60	200.2	10.0	671	1	AI048002 AI048002
61	200.2	10.0	679	7	CN265762 CN265762
62	200.2	10.0	2213	4	CR623485 CR623485
63	200.2	10.0	597	1	AV733437 AV733437
64	200	10.0	598	9	B92813 B92813
65	200	10.0	693	10	AG179363 AG179363
66	199.8	10.0	588	9	AQ581201 AQ581201
67	199.8	10.0	613	9	BZ610480 BZ610480
68	199.8	10.0	624	2	BF854308 BF854308
69	199.8	10.0	670	3	BM671184 BM671184
70	199.8	10.0	681	9	AO543621 AO543621
71	199.8	10.0	909	5	B0881589 B0881589
72	199.8	10.0	528	2	BG005696 BG005696
73	199.6	10.0	624	5	BX508814 BX508814
74	199.6	10.0	501	5	BX485916 BX485916
75	199.4	10.0	561	6	CD688632 CD688632
76	199.4	10.0	712	6	CP146965 CP146965
77	199.4	10.0	849	5	DN602222 DN602222
78	199.4	10.0	933	5	BQ959144 BQ959144
79	199.4	10.0	933	5	BQ959144 BQ959144
80	199.2	10.0	554	7	CN270785 CN270785
81	199.2	10.0	664	10	AG127255 AG127255
82	199.2	10.0	826	5	BO438519 BO438519
83	199.2	10.0	867	5	BO438519 BO438519
84	199.2	10.0	870	9	AQ739907 AQ739907
85	199.2	10.0	924	9	AQ459187 AQ459187
86	199	9.9	516	9	AQ112451 AQ112451
87	199	9.9	541	2	BG682496 BG682496
88	199	9.9	548	1	AI583291 AI583291
89	199	9.9	626	9	AO485636 AO485636
90	199	9.9	740	10	CI983367 CI983367
91	199	9.9	945	5	BU527620 BU527620
92	198.8	9.9	329	1	AI185394 AI185394
93	198.8	9.9	392	1	AI045476 AI045476
94	198.8	9.9	394	9	AO605998 AO605998
95	198.8	9.9	521	10	CG734335 CG734335



96 198.8 9.9 587 9 A0345662 RPEC11-10  
97 198.8 9.9 600 9 A0622755 HS 5339. B  
C 98 198.8 9.9 707 11 CR974543  
C 99 198.8 9.9 844 2 BF694941  
100 198.6 9.9 392 9 AZ254675 AZ254675 UP\_561-7J

## ALIGNMENTS

RESULT 1  
A0398756  
LOCUS  
DEFINITION  
A0398756 494 bp DNA linear GSS 08-JUN-1999  
HS 5338 B1 D03 SPEE RPEC1-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=914 Col=5 Row=H, genomic survey sequence.

ACCESSION  
A0398756  
VERSION  
A0398756.1 GI:50299668  
KEYWORDS  
GSS.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 494)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

REFERENCE  
AUTHORS  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPEC1-11. For BAC  
library availability, please contact Pieter de Jong  
BACPAC Resources (http://bacpac.med.bufileo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 914 row: H column: 5  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 494.  
Location/Qualifiers

## FEATURES

source  
1..494  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=914 Col=5 Row=H"  
/sex="male"  
/clone\_lib="RPEC1-11 Human Male BAC Library"  
/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"

## ORIGIN

Query Match 20.8%; Score 417.2; DB 9; Length 494;  
Best Local Similarity 96.6%; Pred. No. 1.4e-45;  
Matches 425; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1062 AGTAGACCTGGTATTTCTTATACCTTCATTTCAGTAATCAGAAAGATGCTTCATATGCA 1121  
DB 1 AGTATACCTGGTATTTCTTATTCCTTCATTCAGCAATGAGAGATGCTTCATATGCA 60

QY 1122 CAGTGTGTGAGGTACATCAAAAGAAAGAAACAGTTCTTGTTTAAATTTTCAC 1181  
|||||

DB 61 CAGTGTGTGAGGTACATCAAAAGAAAGAAACAGTTCTTGTTTAAATTTTCAC 120  
QY 1182 CGAAAGAAAGAACGCCATTTTGTTCGCTTAATTAAGCAGTCATGACATTAAGAGCC 1241  
DB 121 CGAAAGAAAGAACGCCATTTTGTTCGCTTAATTAAGCAGTCATGACATTAANAGAC 180

QY 1242 AGCAGAGTCCTTTGAAGGCGTGTAACACAGGCTTTTATTATTCACAGCAGACTTGC 1301  
DB 181 ACGCAGAGTCCTTTGAAGGCGTGTAACACAGGCTTTTATTATTCACAGCAGACTTGC 240

QY 1302 ACTTCTACTATGCTTAGAGCTGAAGAAATGGCTCAGAAAGATGAACAATCTCAGAGGCC 1361  
DB 241 ACTTCTACTATGCTTAGAGCTGAAGAAATGGCTCAGAAAGATGAACAATCTCAGAGGCC 300

QY 1362 CTAACTAAGTGAAGCAGGCTGTATTAAGCAACAAGTCAAGAGGTCGAAACTAAGTTC 1421  
DB 301 CTAACTAAGTGAAGCAGGCTGTATTAAGCAACAAGTCAAGAGGTCGAAACTAAGTTC 360

QY 1422 TTGAATCTCCACATCTTTCTTACGTCAGAAAGCCAGCTGATTTTAACTTGAAT 1481  
DB 361 TTGAATCTCCACATCTTTCTTACGTCAGAAAGCCAGCTGATTTTAACTTGAAT 420

QY 1482 TAGAAATTTTAAATTAAT 1501  
DB 421 AGAAATTTTAAATTAATCT 440

RESULT 2  
BZ610799/c  
LOCUS  
DEFINITION  
BZ610799 744 bp DNA linear GSS 08-JUN-2003  
WHAD52TR Human MCF7 breast cancer cell line library (MCF7\_1) Homo  
sapiens genomic clone MCF7\_1-2118, genomic survey sequence.

ACCESSION  
BZ610799  
VERSION  
BZ610799.1 GI:31519360  
KEYWORDS  
GSS.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 744)

REFERENCE  
AUTHORS  
Volik,S., Zhao,S., Chiu,K., Brebner,J.H., Herndon,D.R., Tao,Q.,  
Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,  
Gray,J.W. and Collins,C.  
End-sequence profiling: Sequence-based analysis of aberrant genomes  
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)  
12788976  
Contact: Volik SV  
Colin Collins' Lab  
UCSF Comprehensive Cancer Center  
UCSF Box 0808, San Francisco, CA 94143-0808, USA  
Tel: 415 502 7066  
Fax: 415 502 5665  
Email: svolik@cc.ucsf.edu  
This clone is available from Amplicon Express  
http://www.genomex.com  
Class: BAC ends.  
Location/Qualifiers

## FEATURES

source  
1..744  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="MCF7\_1-2118"  
/sex="female"  
/clone\_lib="Human MCF7 breast cancer cell line library  
(MCF7\_1)"  
/note="Vector: pCBAC1; Site 1: HindIII; This library was  
constructed from MCF7 breast cancer cell line by Amplicon  
Express (http://www.genomex.com) using their standard  
procedure."

## ORIGIN

Query Match 10.7%; Score 214.4; DB 9; Length 744;

Best Local Similarity 73.0%; Pred. No. 6.2e-19;  
Matches 305; Conservative 0; Mismatches 106; Indels 7; Gaps 2;

QY 268 TGTGTTGATGAAATAGATTCACAACTAGTGTGTATTTGAAGATTAAAGCTTA 327  
DB 610 TGTATGAGAAATAGATTTATTTATCAACAGTCTTCAAACTTTAAATTAATGAGG 551  
QY 328 TTTATTACACTATTTAAATTTAAATTTAAATTAATTAATTAATTAATTAAGCTT 387  
DB 550 ATATCATTTCTATGAGACATTTAATAGCAGAAAGTGAAGTAAAGTTAAAGGCTTT 491  
QY 388 TGAATAGGCGCAGGCGAGTGTCTCTGCTGTATTCACAACTTTGGAGCCAAAGT 447  
DB 490 ----CCTGGCCATGCCAGAGGCTTATGCTTTATATCCAGACTTTGGAGGCTGAGG 435  
QY 448 GGGCGGATACCCGAGGTCAGAGCTTAAACAGAGCTGGCCAACTGTGAAACCTCTG 507  
DB 434 GGGGTGATCACCTGAGGTCAGAGTTGAGACAGAGCTGGCCAACTGTGAAACCTCTG 375  
QY 508 CTCTACTAAAAAGCAAAATTTAGCCAGGTGTGTGATGCACTGTATGCCAACTAC 567  
DB 374 CTCTACCAAAAATACAAAATTTAGCTGGTATGTGTGAGGCACTGTATCCAGCTAC 315  
QY 568 TCAGAGGTTGAGGAGGAAATGCTTGAACCTAGAGGTGAGGTTGCACTAACCCGA 627  
DB 314 CTCGGAGGCTGAGGAGGAAATGCTTAACTCAGAGGCGAGGTTGCACTAACCTGA 255  
QY 628 GAT--GTCACTGACTCCAGCTGGCAGACAGCAAGACTCATTAAGCAACAAAA 682  
DB 254 GATCAGCCCTGCACTCCAGCTGGCAGACAGCAAGACTCATTAAGCAACAAAA 197

RESULT 3  
LOCUS BZ603854 871 bp DNA linear GSS 08-JUN-2003  
DEFINITION WHAD093TF Human MCF7 breast cancer cell line library (MCF7.1) Homo  
sapiens genomic clone MCF7\_1-19p17, genomic survey sequence.  
ACCESSION BZ603854  
KEYWORDS BZ603854.1 GI:31512316  
SOURCE GSS.  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 871)  
Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,  
Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,  
Gray,J.W. and Collins,C.  
End-sequence profiling: Sequence-based analysis of aberrant genomes  
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)

TITLE JOURNAL PUBLISHED  
COMMENT CONTACT: Volik SV  
Colin Collins' lab  
UCSF Comprehensive Cancer Center  
UCSF Box 0808, San Francisco, CA 94143-0808, USA  
Tel: 415 502 7066  
Fax: 415 502 5665  
Email: svolik@cc.ucsf.edu  
This clone is available from Amplicon Express  
http://www.genomex.com  
Classes: BAC ends.

FEATURES  
source location/Qualifiers

1..871  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="MCF7\_1-19p17"  
/sex="female"  
/clone\_lib="Human MCF7 breast cancer cell line library  
(MCF7.1)"  
/note="Vector: pECBAC1, Site 1: HindIII; This library was  
constructed from MCF7 breast cancer cell line by Amplicon

Express (http://www.genomex.com) using their standard  
procedure."

Query Match 10.6%; Score 212.8; DB 9; Length 871;  
Best Local Similarity 77.7%; Pred. No. 9.5e-19;  
Matches 283; Conservative 0; Mismatches 77; Indels 4; Gaps 2;

QY 344 AATTAATTTTAAATCACTAATCACTTAATTTAATTAAGCTTTGAATGGCCGAGCG 403  
DB 462 ATTCATATGAGCGGAAATACCTGTGTGTTAAAGCTGTCAAGGCGCAGGCG 521  
QY 404 CAGTACCTCTGCTGTAAATCCCAACACTTTGGAGGCCAAGTGGCGGATCACTGAG 463  
DB 522 CAGTGATACACCTGTATTCCTAGACCTTTGGAGGCCAAGTGGCGGATCACTGAG 581  
QY 464 GTTCAGAGTTTAAAGACAGCTTGGCCCAACATGTGTAAACCTGTCTCTAATAAAGCA 523  
DB 582 GTTCGAGGTTTGAGACCACTGCGCCCAACATGTGTAAACTGTACTTAATAAACACA 641  
QY 524 AAAATAGCAAGTGTGTGTGATGACCTGTATGCTTCCCACTTACAGAGGTTGAGGA 583  
DB 642 AAAATAGCCAGCGCTGTGTGAGACCTGTAAACCCAGCTTACAGAGGCTGAGGA 701  
QY 584 GAGAAATGCTTGAACCTAGAGGAGTGAAGTTGCACTTAACCCAGAGA---TGTCACTGAC 640  
DB 702 AGGAAATGCTTGAACCTAGAGGAGTGAAGTTGCACTTAACCCAGAGA---TGTCACTGAC 761  
QY 641 TCCAGCTT-GGCAAGAGCAAGACTCCATTAAGACAAAGCTTTGAAATTTGTGA 699  
DB 762 TCCAGCTTGGGCAAGAGGCGCAAGACTCCGTCAAAAAAAGCCGCTAAATTTTGAA 821

RESULT 4  
LOCUS CN266085 658 bp mRNA linear EST 16-MAY-2004  
DEFINITION 17000424984279 GRN\_ES Homo sapiens cDNA 5', mRNA sequence.  
ACCESSION CN266085  
KEYWORDS CN266085.1 GI:47282499  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 658)  
Brandenberger,R., Wei,H., Zhang,S., Lei,S., Muraige,J., Fisk,G.J.,  
Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,  
Lebkowski,J. and Stanton,L.W.  
Transcriptome characterization elucidates signaling networks that  
control human ES cell growth and differentiation  
Nat. Biotechnol. 22 (6), 707-716 (2004)

TITLE JOURNAL PUBLISHED  
COMMENT CONTACT: Brandenberger R  
Regenerative Medicine  
Geron Corporation  
230 Constitution Drive, Menlo Park, CA 94025, USA  
Tel: 650 473 8658  
Fax: 650 473 7760  
Email: rbrandenberger@geron.com  
Insert length: 658 Std Error: 0.00.  
Location/Qualifiers

FEATURES  
source

1..658  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="embryonic stem cells, cell lines H1, H7, and  
H9"  
/clone\_lib="GRN\_ES"

## ORIGIN

/note="oligo dt primed, full-length enriched cDNA library from undifferentiated hcs cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions"

Query Match 10.6%; Score 211.4; DB 7; Length 658;  
Best Local Similarity 75.1%; Pred. No. 1.6e-18;  
Matches 277; Conservative 0; Mismatches 91; Indels 1; Gaps 1;

QY 389 GAATGGGCGACGGCGAGTATCCCTGCTTAATCCCAACATTGGGAGGCGAAGGTG 448  
DB 168 GAAAAAGGCGAGGCGGCTGCTTAATCCCTGCTTAATCCCAACATTGGGAGGCGAAGGTG 227  
QY 449 GGGGATCCTCAGGCTGAGGATTAAAGCAGCAGCTGGGCGAACAATGTTAAACCTGTC 508  
DB 228 GGGAAATCAGCTGAGGTTAGGATTGAGCCAGCCTGACCAACATGTTGTAACCCCTGC 287  
QY 509 TCTACTAAAAACGCAAAATTTAGCCAGTGTGTGCAGTACCTGTAGTCCCACTACT 568  
DB 288 TCTACTAAAAATCAAAATTTAGCTGGTGTGATGTAGTACACCTGTGTCCCACTACT 347  
QY 569 CAGGAGCTTGAAGGAGAGAAATGCTTGAACCTTGAAGGAGTGTGAGTAAACCGAG 628  
DB 348 CCGGAGGCTGAGGCGAGAGATGCTTGAACCTTGAAGGAGTGTGAGTAAACCGAG 407  
QY 629 AT-GTCACTGCACTCCAGCTGGCAACAGAGCAAGACTCCATTAAGACAACAAAGCTTT 687  
DB 408 ATCACTATTTACTTCCAGCTGGCGACAGCGAGATCTCGTCTCAAAAAAATTTAAAT 467  
QY 688 GAAATGTGTAAATGAGTTGTACCTTATCTTAAAGAAATTCATCTTTGTCTTAT 747  
DB 468 AAAAGGTAGACGTGGGTCTAGTGACTGTGTGATGATCATCATTAATTAATGAGG 527  
QY 748 TTTTACTGT 756  
DB 528 TTTCTTCTG 536

## RESULT 5

LOCUS BF991881/c 443 bp mRNA linear EST 23-JAN-2001  
DEFINITION PM4-GN0017-291000-004-b03 GN0017 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF991881  
VERSION BF991881.1 GI:12398204  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM4&ct=PM4-GN0017-291000-004-b03&ct3=2000-10-29&ct4=1)

## FEATURES

source

Seq primer: puc 18 forward  
High quality sequence start: 20  
High quality sequence stop: 326.  
Location/Qualifiers

1..443  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="GN0017"  
/note="Organ: Placenta normal; Vector: puc18; Site: 1;  
Sma1; Site: 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 10.4%; Score 208.8; DB 2; Length 443;  
Best Local Similarity 81.0%; Pred. No. 4.1e-18;  
Matches 256; Conservative 0; Mismatches 57; Indels 3; Gaps 1;

QY 369 TTTAATTTAATAAGACTTTGAATGGCCAGCGCAGTAGCTCCCTGTAATCCCA 428  
DB 414 TCAAACTTAAAAAGAGGCTGGGTGTGCTGAGGCTGATGCTGTAATCCAG 355  
QY 429 CACTTGGGAGGCGAAGTGTGGCGATCACTGAGTCAAGATTAAAGCAGCCTGGC 488  
DB 354 CACTTGGGAGGCGAAGTGTGGCGATCACTGAGTCAAGATTAAAGCAGCCTGGC 295  
QY 489 CAACATGTGAAACCTGTCTCTACTTAAAGCAAAATTTAGCCAGTGTGTGCATG 548  
DB 294 CAACATGTGAAACCTGTCTCTACTTAAATTCATAAATTTAGCTGGCGTGGTCGCA 235  
QY 549 CACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATTCCTTGAACCTAGAGGT 608  
DB 234 TGCTGTAAATCCAGCTACTCGGAGGCTGAGGCGAGAAATTCCTTGAACCGAGGAT 175  
QY 609 GAGGTTGAGTACAGCAGGAT--GTCACTGCACTCCAGCTGGCAAGCAGAGACT 665  
DB 174 GAGGTTGAGTACAGCAGGATCAACCACTGCACTCCAGCTGGCAAGCAGAGACT 115  
QY 666 CCATTAAGACAACAAA 681  
DB 114 CCCTTTAAAAACAAA 99

## RESULT 6

LOCUS BF991882/c 435 bp mRNA linear EST 23-JAN-2001  
DEFINITION PM4-GN0017-291000-004-b04 GN0017 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF991882  
VERSION BF991882.1 GI:12398205  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics

Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and

201203, P. R. China





REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
1 (bases 1 to 812)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Place: LBL0075 row: g column: 02  
High quality sequence scop: 655.  
Location/Qualifiers

FEATURES  
source

1..812  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:438761"  
/issue\_type="osteosarcoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 86"  
/note="Organ: bone; Vector: pCMV-Sport6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.533 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 10.3%; Score 206; DB 2; Length 812;  
Best Local Similarity 85.0%; Pred. No. 7,6e-18;  
Matches 243; Conservative 0; Mismatches 40; Indels 3; Gaps 1;

QY 400 GGGGCGTGTCTCTGCTGCTGTAATCCCAACATTTGGGAGCCAGGTGGCGGATCACC 459  
DB 242 GGTGCGTGTCTGCTGCTGTAATCCCAACATTTGGGAGCCAGGTGGCGGATCACC 301  
QY 460 TGAAGTCAGAGATTAAAGACAGCCTGGCCACATGTTGAAACCTGTCTCTAATAAA 519  
DB 302 TGAAGTCAGAGATTAAAGACAGCCTGGCCACATGTTGAAACCTGTCTCTAATAAA 361  
QY 520 CGCAAAATTAAGCAGGTGTGTGTCGATGCACTGTATCCCACTACTCAGAGGTTGA 579  
DB 362 TACAAAAATTAAGCAGGTGTGTGTCGATGCACTGTATCCCACTACTCAGAGGTTGA 421  
QY 580 GGGAGGAGATTCCTTGAACCTAGAGATGAGGTTGAGTAACCCGAGATGTC--ACT 636  
DB 422 GGCACGAGATTCCTTGAACCCAGAGGCGAGGTTGAGTAACCCGAGATGTC--ACT 481  
QY 637 GCACTCCAGCCTGGCAACAGAGCAAGATCCATTAAGCAACAATAA 682  
DB 482 GCACTCCAGCCTGGCAACAGAGCAAGATCCATTAAGCAACAATAA 527

RESULT 13  
BUB49843 710 bp mRNA linear EST 16-OCT-2002  
LOCUS BUB49843  
DEFINITION AGENCOURT 10440376 NIH\_MGC\_109 Homo sapiens cDNA clone  
IMAGE:6598537 5', mRNA sequence.  
ACCESSION BUB49843  
VERSION BUB49843.1 GI:24034806  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
REFERENCE 1 (bases 1 to 710)  
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE  
JOURNAL  
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Place: LBL02811 row: i column: 01  
High quality sequence scop: 592.  
Location/Qualifiers

FEATURES  
source

1..710  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6598537"  
/issue\_type="teratocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 109"  
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:  
XhoI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

## ORIGIN

Query Match 10.3%; Score 205.6; DB 5; Length 710;  
Best Local Similarity 75.7%; Pred. No. 9,1e-18;  
Matches 281; Conservative 0; Mismatches 85; Indels 5; Gaps 2;

QY 317 TTTAAGATTATTTATTTACACTATTATTAATAATTTTAAACCTATACCTTAATTA 376  
DB 333 TTCAGGTATATATATATATATATATAGTAAGTAATTAAGTAATTAATTTACGTTACTTT 392  
QY 377 TTTAAGATTATTTAATTTGGCCAGGCGAGTGTCTGCTGCTGTAATCCCAACACTTTGG 436  
DB 393 TTTAAGATTATTTAATTTGGCCAGGCGAGTGTCTGCTGCTGTAATCCCAACACTTTGG 452  
QY 437 GAGCCAGAGTGGCGAGTACCTGAGTCAAGATTTTAAAGACAGCCTGGCCACATGG 496  
DB 453 GAGCCAGAGTGGCGAGTACCTGAGTCAAGATTTTAAAGACAGCCTGGCCACATGG 512  
QY 497 TGAACCTGTCTCTACTAATAAAGCAAAATTAAGCAGGTGTGTGTCATGCACTGTA 556  
DB 513 TGAACCTGTCTCTACTAATAAATTAAGCAGGTGTGTGTCATGCACTGTA 572  
QY 557 GTTCCCACTTCTCAGAGGTTGGAGGAGATTTGCTTGAACCTTGAAGAGTGAAGTTGA 616  
DB 573 ATCCCACTTCTCAGAGGTTGGAGGAGATTTGCTTGAACCTTGAAGAGTGAAGTTGA 632  
QY 617 CAGTAACCCGAGATGTAC--TGCACTCCAGCCT--GGCAACAGAGCAAGACTTCATAA 671  
DB 633 CAGTAACCCGAGATGTAC--TGCACTCCAGCCT--GGCAACAGAGCAAGACTTCATAA 692  
QY 672 AGACAACAATAA 682  
DB 693 CAAAAAATAA 703

RESULT 14  
AUI47414 544 bp mRNA linear EST 05-AUG-2002  
LOCUS AUI47414/c  
DEFINITION AUI47414 MAMMAL Homo sapiens cDNA clone MAMMAL1000605 3', mRNA  
sequence.  
ACCESSION AUI47414  
VERSION AUI47414.1 GI:11008935  
KEYWORDS EST.



SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 544)  
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Makamatsu, A., Ozawa, M., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.).  
TITLE HRI human cDNA project (Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Makamatsu, A., Ozawa, M., Nakamura, Y., Nagai, T., Sugano, S., Isogai, T.)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
LOCATION/Qualifiers  
1. 544  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MAMMA100605"  
/issue\_type="mammary gland"  
/clone\_lib="MAMMA1"  
/note="Vector: pME18SFL3"

ORIGIN  
Query Match 10.3%; Score 205.4; DB 1; Length 544;  
Best Local Similarity 85.7%; Pred. No. 1.1e-17;  
Matches 240; Conservative 0; Mismatches 37; Indels 3; Gaps 1;  
QY 390 AATGGGCGAGCGCAGTAGCTCTGCTGTAAATCCCACTTTGGAGGCCAAGGTG 449  
DB 288 ATGAGAGCGCAGGCGGTGGCTCACACCTGCAATCCAGCACTTGGAGGCCAAGGTG 229  
QY 450 GGGAGTACCTGAGGTGAGAGTTTAAGACCAAGCTGGCCAAATGTAACCTGTCT 509  
DB 228 GCGAGTACCTGAGGTGAGAGTTGAAGACCAAGCTGGCCAAATGTAACCTGTCT 169  
QY 510 CTACTAAAAACGCAAAATTTAGCCAGGTGAGGATGACCTGTAGTCCCAACTACTC 569  
DB 168 CTACTAAAAACGCAAAATTTAGCCAGGTGAGGATGAGGATGAGTCCCAACTACTC 109  
QY 570 AGGAGGTGAGGAGGAGAAATTTGAACTGAGAGGTGAGGAGTTGCAAGTAAACCGAG 628  
DB 108 AGGAGGTGAGGAGGAGAAATTTGAACTGAGAGGTGAGGAGTTGCAAGTAAACCGAG 49  
QY 629 --ATGTCATGCACTCCAGCTGGCAACAGAGCAACTC 666  
DB 48 TCATGCCACTGTACTCCAGCTGGCAATGAGCTAATTC 9

RESULT 15  
LOCUS CR936687 2768 bp mRNA linear HTC 23-FEB-2005  
DEFINITION Homo sapiens mRNA; cDNA DKFZp686H23233 (from clone DKFZp686H23233).  
ACCESSION CR936687  
VERSION CR936687.1 GI:60219598  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 2768)

AUTHORS Bahr, A., Lauber, J., Mewes, H.W., Well, B., Amid, C., Oeanger, A., Fob, G., Han, M. and Wiemann, S.  
CONSTRM The German cDNA Consortium  
TITLE Direct Submission  
JOURNAL Submitted (22-FEB-2005) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY  
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.  
This clone (DKFZp686H23233) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.  
Please contact RZPD for ordering:  
http://www.rzpd.de/cgi-bin/products/ci.cgi?cloneid=DKFZp686H23233  
Further information about the clone and the sequencing project is available at <http://mips.gsf.de/projects/cdna/>.  
LOCATION/Qualifiers  
1. 2768  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp686H23233"  
/issue\_type="small intestine"  
/clone\_lib="686 (synonym: hlcc3). Vector pSport1\_SfiI, host DH10B; sites SfiIA + SfiIB"  
/dev\_stage="adult"

ORIGIN  
Query Match 10.3%; Score 205.2; DB 4; Length 2768;  
Best Local Similarity 80.6%; Pred. No. 6.2e-18;  
Matches 253; Conservative 0; Mismatches 58; Indels 3; Gaps 1;  
QY 372 AATTATTAAGAGCTTTGAATGGCCAGCGCAGTAGCTCTGCTGTATCCCAAC 431  
DB 2447 AATGAAAAAAATATTTGGCTTGGCCAGCGCGGTGCTCACGCTGTATATCCAGCAG 2506  
QY 432 TTTGGAGGCCAAGAGTGGGCGATGACCTGAGGTGAGGATTAAACCAAGCTGGCCAA 491  
DB 2507 TTTGGAGGCCAAGAGTGGGCGATGACCTGAGGTGAGGATTAAACCAAGCTGGCCAA 2566  
QY 492 CATGTGAAACCTGTCTCTACTTAAACGCAAAATTTAGCCAGGTGAGTGGCATGCAC 551  
DB 2567 CATGTGAAACCTGTCTCTACTTAAACGCAAAATTTAGCCAGGTGAGTGGCATGCAC 2626  
QY 552 CTGTAGTCCCAACTACTCAAGAGTTGAGGAGAGAAATTTGTAACCTAGAGGTGGA 611  
DB 2627 CTGTAGTCCCAACTACTCAAGAGTTGAGGAGAGAAATTTGTAACCTAGAGGTGGA 2686  
QY 612 GGTTCAGTAACCCGAGAT--GTCACTGACCTCCAGCTGGCAACAGAGCAAGTCCCA 668  
DB 2687 GGTTCAGTAACCCGAGAT--GTCACTGACCTCCAGCTGGCAACAGAGCAAGTCCCA 2746  
QY 669 TAAAGACACAAAA 682  
DB 2747 TTTCAAAAAAAA 2760

RESULT 16  
LOCUS A0037381 455 bp DNA linear GSS 11-JUL-1998  
DEFINITION CIT-HSP-2335120.TR CIT-HSP Homo sapiens genomic clone 2335120, genomic survey sequence.  
ACCESSION A0037381  
VERSION A0037381  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 455)  
Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Baes, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,

TITLE Simon, M. and Venter, J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map

JOURNAL Building (1998)  
COMMENT Unpublished (1998)  
Other GSSs: CIT-HSP-2335L20.TF  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page: [http://www.tigr.org/cdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES  
source Location/Qualifiers

```
1..455
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2335L20"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/notes="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"
```

## ORIGIN

```
Query Match      10.2%; Score 205; DB 9; Length 455;
Best Local Similarity 80.4%; Pred. No. 1.3e-17;
Matches 286; Conservative 0; Mismatches 60; Indels 5; Gaps 2;

QY 356 AAAAATAATACACTTAATTATTAAAGAGCTTTGAATGGCCAGAGCGAGTAGCTCTG 415
DB 98 AAAAATTAATAAAGTGTCTTTAAACAGACTCACACATAGCCAGAGCGAGTGGCTCATG 157
QY 416 CCTGTAAATCCCAACACTTTTGGAGAGCCAGAGTGGCGGATCACTGAGGTCCAGAGTTTA 475
DB 158 CCTGTATATCCAGACTTTTGGAGAGCCAGAGCGGAGTGGATCACTTGAAGTCCAGAGTTCA 217
QY 476 AGACCAAGCCTGGCCAAATGATGTAACCTGTCTCTACTAAAAACGAAAAATTAGCCAG 535
DB 218 AGACCAAGCCTGGCCAAATGATGTAACCTGTCTCTACTAAAAATTAAGCCAG 277
QY 536 GTGTGTGATGATGACCTGTAGTCCCACTACTCAGAGAGTTGAGGAGGAATTTGCTT 595
DB 278 GCGTGTGTGGGGGGGCGCTGTATATCCAGCTACTCAGAGAGGCTGAGGAGGAATCGCTT 337
QY 596 GAACTTAGAGAGTGGAGTTGACATAACCCGAGA---TGTCACTGCACTCCAGCTCG-- 650
DB 338 GAACTTAGAGAGTGGAGTTGACATAACCCGAGA---TGTCACTGCACTCCAGCTCG-- 397
QY 651 CAACAGAGCAAGACTCCATAAAGACAACAA 681
DB 398 TACAAGAGCGAAACTCATCTCAAAACAAA 428
```

RESULT 17  
CD699526 462 bp mRNA linear EST 25-JUN-2003  
LOCUS CD699526  
DEFINITION EST16162 human nasopharynx Homo sapiens cDNA, mRNA sequence.  
ACCESSION CD699526  
VERSION CD699526.1 GI:32229113  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 462)  
AUTHORS Liu, X.-Q., Zhou, Y., Zhang, L.-J., Xu, H., Chen, H.-K., Fan, Z.-G. and

TITLE Zeng, Y.-X.  
Transcriptional Gene Expression Profile of Human Nasopharynx  
JOURNAL Unpublished (2003)  
COMMENT Contact: Yixin Zeng  
Cancer Center  
Sun Yat-sen University  
651 Dongfeng Road East, Guangzhou 510060, China  
Tel: 86-1380-9770-743  
Fax: 86-20-8775-4506  
Email: yxzeng@zsunm.edu.cn

FEATURES  
source Location/Qualifiers

```
1..462
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/notes="ESTs generated from a normal nasopharynx cDNA library from Southern Chinese"
```

## ORIGIN

```
Query Match      10.2%; Score 205; DB 6; Length 462;
Best Local Similarity 84.1%; Pred. No. 1.3e-17;
Matches 243; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

QY 395 GGCAGAGCGCAGTAGTCTCTGCTGATCCCAACACTTTGGAGGCCAAGTGGCGGA 454
DB 34 GGCAGAGCGCGGGGGGCTCATGCTGCAATCCAGACTTTGGAGGCCAGAGTGGCGGA 93
QY 455 TCACCTGAGGTGAGAGTTTAAGACCAAGCTGCGCCAAACATGGTGAACCTGCTCTACT 514
DB 94 TCACCTGAGGTGAGAGTTTAAGACCAAGCTGCGCCAAACATGGTGAACCTGCTCTACT 153
QY 515 AAAAAGCAAAATTAAGCAGAGTGTGGATGACCTGTAGTCCCACTACTCAGAG 574
DB 154 AGAAATACAAATTAAGCAGAGTGTGGATGACCTGTAGTCCCACTACTCAGAG 213
QY 575 GTTGAAGAGAGGAATTTGCTTAACCTAGAGGTGAGGTGCAATACCCAGATGTCA 634
DB 214 GCTGAGGAGAGGAATCGTTTGAAGCCCTGGAAGCAGAGTTGCAATGAGTGAAGAGCA 273
QY 635 CTGCACTCCAGCCT-GGCAACAGAGCAAGACTTCATTAAGACAACAAA 682
DB 274 CTGCACTCCAGCCTGGGTGAGCAGAGCAAGCTTCGTCACAAAAAAA 322
```

RESULT 18  
BZ612424 693 bp DNA linear GSS 08-JUN-2003  
LOCUS BZ612424  
DEFINITION WHADW75TR Human MCF7 breast cancer cell line library (MCF7\_1) Homo  
sapiens genomic clone MCF7\_1-22M5, genomic survey sequence.  
ACCESSION BZ612424  
VERSION BZ612424.1 GI:31520985  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 693)  
AUTHORS Volik, S., Zhao, S., Chin, K., Brebner, J. H., Herndon, D. R., Tao, Q.,  
Kowbel, D., Huang, G., Lapuk, A., Kuo, W.-L., Magrane, G., de Jong, P.,  
Gray, J. W. and Collins, C.  
End-sequence profiling: Sequence-based analysis of aberrant genomes  
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)  
JOURNAL PUBMED  
COMMENT Contact: Volik SV  
Colin Collins' lab  
UCSF Comprehensive Cancer Center  
UCSF Box 0808, San Francisco, CA 94143-0808, USA  
Tel: 415 502 7066  
Fax: 415 502 5665  
Email: svolik@cc.ucsf.edu





This clone is available royalty-free through LNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information. This read is a RESEQUENCE of a previously sequenced human clone. Original clone citation: Mashu-Merck EST Project. This read has been verified (found to hit its original self in the correct orientation).

Insert Length: 1908 Std Error: 0.00  
Seq primer: -40RP from Gibco  
High quality sequence stop: 381.

## FEATURES

source

```
1..458
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:4619634"
/db_xref="taxon:9606"
/clone="IMAGE:587332"
/cisue_type="tumor"
/cell_line="T84 carcinoma cell line"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Stratagene colon (#937204)"
/notes="Organ: colon; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt-T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTATTTTATTTT 3' "
```

## ORIGIN

Query Match 10.2%; Score 204.2; DB 1; Length 458;  
Best Local Similarity 75.3%; Pred. No. 1.6e-17;  
Matches 281; Conservative 0; Mismatches 88; Indels 4; Gaps 2;

```
QY 314 AGATTAAAGTATTATTATTAACAATATTAAATTTTAAACTAATACCTTAA 373
DB 398 AGTTATAGACATTGTGGTTAACTTTGGAGTCCAAAGCTACTCTAAGAGAAATTAAT 339
QY 374 TATTAAAGAGCTTTGAAATGGGCGAGGCGAGTACCTCGCTGTAATCCCAACATT 433
DB 338 AATATATCTTTTATTTTGGGCGAGGCGGTGTGCTATGCTTAAATCCAGACATT 279
QY 434 TGGGAGGCGCAAGGTGGGCGGATCACTGAGTCAAGAGTTTAAAGACAGCCTGGCCAA 493
DB 278 TGGGAGTCCGAGGCGAGTGGATTCACCTGAGTCAAGAGTTCCAGACCAAGCTGACCA 219
QY 494 TGGTAAACCTCTCTCTACTTAAACCGAAATTTAGCAGGTGTGTGTCATGCACT 553
DB 218 TGGTAAACCTCTCTCTCTACTTAAACCGAAATTTAGCAGGTGTGTGTCATGCACT 159
QY 554 GTAGTCCCAACTACTCAGAGGTTGAGGAGGAGAAATGCTTGAACCTTGAAGTGAAG 613
DB 158 ATATATCCAGCTACTCAGAGGCTGAGGCGAGGAGATCGCTTGAACCTTGAAGTGAAG 99
QY 614 TTGCACTGAACCCGAGA--TGTCACTGCACTCCAGCT--GGCAACAGAGCAAGCTCCT 669
DB 98 TTGCACTGAGCGAGATCGTGCACATGCACTCCAGCTCGGTGTGAGAGCAAGATCTCCAT 39
QY 670 AAAGACAAACAAA 682
DB 38 CTCAAAAA 26
```

RESULT 23  
AA808875/c 390 bp mRNA linear EST 18-FEB-1998  
DEFINITION nt8h02.s1 NCI\_CGAP\_P33 Homo sapiens cDNA clone IMAGE:1205427  
similar to contigins Alu repetitive element; mRNA sequence.  
AA808875  
VERSION AA808875.1 GI:2878281  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE  
1 (bases 1 to 390)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

## JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb@email.nih.gov](mailto:cgapsb@email.nih.gov)  
Tissue Procurement: W. Maxson Linehan, M.D., Rodrigo Chuquai,  
M.D., Michael Emmert-Buck, M.D., Ph.D.,  
CDNA Library Preparation: David B. Krizman, Ph.D.  
DNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 929 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 344.

## FEATURES

source

```
1..390
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1205427"
/dev_stage="45 years old"
/lab_host="DH10B"
/clone_lib="NCI CGAP_P33"
/notes="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA was primed with oligo(dt)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected cells histologically-determined to be fully malignant prostate cancer cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDE-cloning method (life technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."
```

## ORIGIN

Query Match 10.2%; Score 203.6; DB 1; Length 390;  
Best Local Similarity 80.3%; Pred. No. 2.1e-17;  
Matches 252; Conservative 0; Mismatches 59; Indels 3; Gaps 1;

```
QY 372 AATTATTAAGAGCTTTGAAATGGGCGAGGCGAGTACCTCGCTGTAATCCCAAC 431
DB 326 AATGAAAAAATATTTGGCTTGGCGAGGCGGCGGTGCTCAAGCTTGAATCCAGCAC 267
QY 432 TTTGGAGGCGCAAGGTGGGCGGATCACTGAGTCAAGAGTTTAAAGCAAGCCTGGCCAA 491
DB 266 TTTGGAGGCGCAAGGTGGGCGGATCACTGAGTCAAGAGTTTAAAGCAAGCCTGGCCAA 207
QY 492 CATGGTGAACCCCTGCTTACTTAAACCGCAAAATTTAGCAGGTGTGTGTCATGCA 551
DB 206 CATGGTGAACCCCAATCTACTTAAATAAATAATTAAGCGGGGTGTGTGTATGAC 147
QY 552 CTGTAGTCCCACTACTCAGAGGTTGAGGAGGAGAAATTTGCTGAACCTTGAAGGTGA 611
DB 146 CTGTATATCCCACTACTCGGAGGCTGAGGAGGAGAAATTTGCTTGAACCAAGAGGGGA 87
QY 612 GGTTCAGTAAACCCAGAT--GTCACTGCACTCCAGCCTGGCAACAGAGCAAGACTCCA 668
DB 86 GGTTCAGTAAACCCAGATCAAGCATATGCACTCCAGCCTGGGTAAACAGAGAAATCTGA 27
QY 669 TAAAGCAACAAA 682
DB 26 TTTCAAAAAA 13
```

RESULT 24

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catherhini; Homidae; Homo.
REFERENCE	1 (bases 1 to 545)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBLISHED	10737800
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=MR2&t2=MR2-EN0092-261200-001-e1&t3=2000-12-26&t4=1) Seq primer: puc 18 Forward High quality sequence start: 22 High quality sequence stop: 532. Location/Qualifiers 1..545 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_lib="EN0092" /note="Organ: lung normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
FEATURES	
source	
ORIGIN	
Query Match	10.2%; Score 203.4; DB 2; Length 545;
Best Local Similarity	80.9%; Pred. No. 1.9e-17;
Matches 250; Conservative	0; Mismatches 56; Indels 3; Gaps 1.
OY	377 TTTAAGAGCTTTGAAATGCGGCGAGCGAGTACGCTCCCTGTTATGCCAACTTTGG 436
Db	206 TTTTAAGGCAATTATGTTTGGCCAGCGGAGTGTTCACACCTATATATCCAGCAATGG 265
OY	437 GAGGCAAGGTGGGGGATCACTGAGGTCAAGATTTAAGACCAAGCTGGCCAACTATG 496
Db	266 GTGGCAAGGTGGGCAATCACTTGAGTCAAGAGTTGAGACCAAGCTGGCCAACTATG 325
OY	497 TGAACCCCTGTCTCTACTTAAAAAGCAAAAAATTAGCCAGGTGTGGTCATGCACTGTA 556
Db	326 TGAACCCCACTCTACTTAAAAATACAAAAATTAGCCAGGATGTGGCGTGCACCTGTA 385
OY	557 GTCCCACTACTCAGAGAGTGGAGGAGAGAAATTGCTTGAACCTTAGAGGTGGAGTTG 616
Db	386 ATTCCAGTGTCTTGGAGAGCTGAGGCAAGAAATGCTTGAACCTGGAGAGGGGAGATTG 445
OY	617 CAGTAAACCCGAGA---TGTCACGTGACATCCAGCCGCGGCAACAGACCAAGATCCATAAG 673
Db	446 CAGTGAAGCAAGATTGTGCACGTCCAGCCTGTGTGACAGATGAGATCTCACTCA 505
OY	674 ACAACAAA 682

DB	506	AAAAAAAA	514
RESULT 26			
CR961104			
LOCUS			
DEFINITION	CR961104	526 bp	DNA linear GSS 06-JUN-2005
	Homo sapiens BAC end sequence of RZPDB737G012026D from genomic library (orig. Pieter J. de Jong library RPCT-11), genomic survey sequence.		
ACCESSION	CR961104		
VERSION	CR961104.1	GI:66991670	
KEYWORDS	GSS.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.		
AUTHORS	1 (bases 1 to 526)		
TITLE	Schwarz, F., Neubert, P., Schneider, D., Peters, M. and Korn, B.		
JOURNAL	Direct Submission		
	Submitted (05-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Im Neuenheimer Feld 515, D-69120 Heidelberg, Germany		
COMMENT	RZPD: RZPDB737G012026D: RP11-69M1; derived from Pieter J. de Jong library RPCT-11; http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPDB737G012026D RZPDLIB: (Human Genomic Set - RZPD 1.0) RZPD LIB No. 737 http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737 http://www.rzpd.de/products/genomicaet/Contact: Inge Arlart RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Heubnerweg 6, D-14059 Berlin, Germany Tel: +49 30 32639 100 Fax: +49 30 32639 111 www.rzpd.de This clone is available from RZPD; Contact RZPD (customer.service@rzpd.de) for further information. Clone distribution: http://www.rzpd.de/products/genomicaet/Seq-primers: T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG) Class: BAC ends.		
FEATURES			
source	location/Qualifiers		
	1..526		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/clone="RZPDB737G012026D"		
	/sex="Male"		
	/clone_lib="RPCT-11"		
	/note="Vector: pBAC3.6; RPCT-11 Human Male BAC Library"		
ORIGIN			
Query Match	10.1%;	Score 203;	DB 11; Length 526;
Best Local Similarity	74.4%;	Pred. No. 12e-17;	
Matches 270;	Conservative 0;	Mismatches 90;	Indels 3; Gaps 1;
QY	380 AAGAGCTTTGAATGGGCGCAGGCCAGTAGCTCTGCTGCTGATATCCAACTTTGGAG	439	
DB	106 AAGATTTAAGACGAGGCGCGCGGATTCACACTGTAAATCCCAACTTTGGAG	165	
QY	440 GCCAAGTGGGCGGATCACTGAGGTCAGAGTTTAAAGACCAAGCTGGCCAACTGTGA	499	
DB	166 GCTAGAGGTGGATCACTGAGGTCAAGAGTTTAAAGACCAAGCTGGCCAACTGTGA	225	
QY	500 AACCTGCTCTACTATAAAACGCCAAATTTGCGACAGGTGGTGGCATGACCTGTAGTC	559	
DB	226 AACTGTGCTCTACTATAAAATTAAGATTTGTTGGGATGTGTGCACATGCTGTAGTC	285	
QY	560 CCAACTACTCAGAGGTTAGGAGGAGCAATTTGCTTGAACCTAGAGGTGAGGTTGAC	619	
DB	286 CCAAGTACTCGGAGGCTGAGGAGGAGAAATAGCTTGAACCGGAGGACAGAGGTTGAC	345	
QY	620 TAACCGAG--ATGTCACTGCACTCCAGCTGGCAACAGCAAGATCCATTAAAGCA	676	

Db	346	TGAGCCGAGGACGACCGACTGCACACTCCAGCCTGGCAACAGAGTGAAGCTCTGTCTCAAA	405
Qy	677	ACAAAAGCTTTGAATAATGTGTAAATGATGTTGACCTATCTTCATTTAAGAAATTCATCTT	736
Db	406	AAATTAATAATACATTGATTCAGTGTGATTTACTGTATACCTAGTTGAACCTATAGATATGAGTTAT	465
Qy	737	TGT 739	
Db	466	GGT 468	
RESULT 27			
CR859576/c	CR859576	3165 bp	mRNA
DEFINITION	Pongo pygmaeus mRNA; cDNA DKFPz468A085 (from clone DKFPz468A085).		
ACCESSION	CR859576		
VERSION	CR859576.1	GI:55730029	
KEYWORDS	HTC.		
SOURCE	Pongo pygmaeus (orangutan)		
ORGANISM	Pongo pygmaeus		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pongo.		
AUTHORS	1 (bases 1 to 3165) Ansojge,W., Krieger,S., Regiert,T., Ritzmuller,C., Schwager,B., Mewes,H.W., Well,B., Amid,C., Oeanger,A., Fobo,G., Han,M. and Wiemann,S.		
CONSRPTM	The German CDNA Consortium		
TITLE	Direct Submission		
JOURNAL	Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY		
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFPz468A085) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneid=DKFPz468A085 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.		
FEATURES	Location/Qualifiers		
source	1..3165		
	/organism="Pongo pygmaeus"		
	/mol_type="mRNA"		
	/db_xref="taxon:9600"		
	/clone="DKFPz468A085"		
	/tissue_type="heart"		
	/clone_id="468 (synonym: phrt1). Vector pSport1_Sfi, host DH10B; sites SfiIa + SfiIb"		
	/dev_stage="adult"		
	/note="hypothetical protein (Homo sapiens), differentially spliced"		
gene	1..3165		
	/gene="DKFPz468A085"		
CDS	34..687		
	/gene="DKFPz468A085"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="CAH91740.1"		
	/db_xref="GI:55730030"		
	/translation="MKRVNSCVKSDHVLBELETGEROLKSLLOHLDTSVIERCM SKESPTKLRATPEAIRNLODIDBERIYSERORTICLORPAKSKOLTRRMEIETKSLFO GAORHSFLKALYYOAYHKTSADVKTSIMRKIKIGTKALLVCSOTAMKKYRWLNF"		
ORIGIN	Query Match 10.1%; Score 203; DB 4; Length 3165;		
	Best Local Similarity 74.6%; Pred. No. 1,le-17;		
Matches 296;	Conservative 0; Mismatches 95; Indels 6; Gaps 3;		
Qy	292	AAACAAGTAGCTGCTATTGAAGATTTAAGACTTATTTATTAACACTATTAATAAATAAT	351



Db 3125 AAATGAGATTGCTATGTAGAAACTCTAGATATCATTAATAAACTTTAGAAATATAA 3066  
 QY 332 TTTAAATACTATACATTAATTAATTAAGAGCTTTGAA--ATGGCCAGGCGAGTAG 409  
 Db 3065 GAGTTAGTCAAGAGATACAAATTAATATAGAAAAATCACTTAGCCAGACAGCTGG 3006  
 QY 410 CTCCTGCTGTAATCCCAACACTTTGGAGGCCAAGTAGGCGGATCAGCTAGGTCAAG 469  
 Db 3005 CTCATGCTGTATATCCCACTATATGGAGGCTGAAGAGGAGATCAGCTTAGGTCAAG 2946  
 QY 470 AGTTTAAGACCAAGCCTGCGCAACATGTGAAACCTGTCTTAATAAAACGCAAAATT 529  
 Db 2945 AGTTTGAAGACAGCTTAGCCAAATATGTGAAACCTGTCTTAATAAAATCAAAATT 2886  
 QY 530 AGCCAGGTGTGGTGGATCAGCTGATAGTCCCACTACTCAGAGAGTTAGAGAGAGAA 589  
 Db 2885 AGCCAGGCGCTGTGGGCGCATGCTGTATATCCAGCTACTCAGAGGCTGAGGAGAGAA 2826  
 QY 590 TTGCTTGAACCTAGAGGTGAGAGTTGACAGTAACCCGAGA--TGTCACTGCACTCCAG 646  
 Db 2825 TGGCTTGAACCTAGAGGTGAGAGTTGCACTGAGCCGAGATAGTGCCACTGCACTCCAG 2766  
 QY 647 CT-GGCCACAGACAGACTCCATMAAGCAACAA 682  
 Db 2765 CTGGGTGACAAAGCAAGATCCATTAATAAAAAA 2729

RESULT 28  
 CD701193 492 bp mRNA linear EST 25-JUN-2003  
 LOCUS EST17717 human nasopharynx Homo sapiens cDNA, mRNA sequence.  
 ACCESSION CD701193  
 VERSION CD701193.1 GI:32231823  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 492)  
 Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and Zeng,Y.-X.  
 Transcriptional Gene Expression Profile of Human Nasopharynx  
 Unpublished (2003)  
 TITLE JOURNAL  
 COMMENT Contact: Yixin Zeng  
 Cancer Center  
 Sun Yat-sen University  
 651 Dongfeng Road East, Guangzhou 510060, China  
 Tel: 86-1380-9770-743  
 Fax: 86-20-8775-4506  
 Email: yxzeng@zsums.edu.cn.

FEATURES  
 source  
 1..492  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="normal nasopharynx"  
 /clone\_lib="human nasopharynx"  
 /note="ESTs generated from a normal nasopharynx cDNA library from southern Chinese"

## ORIGIN

Query Match 10.1%; Score 202.8; DB 6; Length 492;  
 Best Local Similarity 87.2%; Pred. No. 2.4e-17;  
 Matches 246; Conservative 0; Mismatches 32; Indels 4; Gaps 2;  
 QY 392 ATGGCCAGGCGCAGTAGTCTCTGCTGATATCCCAACACTTTGGAGGCCAAGGTGGGC 451  
 Db 203 AGGACCAAGGCGCATGAGCTCAGCCCTGTATTCAGACACTTGTGGAGGCTGAGGTGGGC 262  
 QY 452 GGAATCACCTGAGGTGAGAGTTTAAGACAGAGCTGGCCACATGATGTAACCTGTCTCT 511  
 Db 218 TCACTGAGGTGAGAGTTTAAGAACAGCTGTGATCAATGATGTAACCTCATCTACT 159

Db 263 GGATCAGCTGAGGTGAGAGTTTCGAGACCAAGCTGGCCAGCATGTGAAACCTGTCTCT 322  
 QY 512 ACTAAAAAGCAAAATTAAGCCAGGTGTGTGGCAATGACCTGATGCCAATCACTAG 571  
 Db 323 ACTAAAAATACAAAATTAAGCCAGAGGTGTGGCAAGCAGCTGTATCCAGCTACTAG 382  
 QY 572 GAGTTGAGGAGAGAAATTGCTTGAACCTTAGAGAGTGTGAGTTGACTTAACCCAG-- 628  
 Db 383 GAGGCTGAGGCGAGGAATCGCTTGAACTGGAGAGCAGAGGTTTCAGTGTAGCCAAATC 442  
 QY 629 ATGTGACTGCACTCCAGCTGG-CAACAGAGCAAGACTCCAT 669  
 Db 443 ATGCCACTGCACTCCAGCTGGCCAGCCGAGCAGACTCCAT 484

RESULT 29  
 BU957691/c 888 bp mRNA linear EST 21-OCT-2002  
 LOCUS AGENCOURT\_10620381 NIH\_MGC\_107 Homo sapiens cDNA clone  
 DEFINITION IMAGE:6731767 5', mRNA sequence.  
 ACCESSION BU957691  
 VERSION BU957691.1 GI:24187263  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 888)  
 NIH-MGC <http://mhc.nci.nih.gov/>,  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 TITLE JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-r@mail.nih.gov  
 Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 plate: LLCM3058 row: 1 column: 06  
 High quality sequence stop: 610.

FEATURES  
 source  
 1..888  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6731767"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 107"  
 /note="Organ: breast; Vector: pDRB7; Site 1: EcoRI; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 10.1%; Score 202.8; DB 5; Length 888;  
 Best Local Similarity 85.6%; Pred. No. 2e-17;  
 Matches 238; Conservative 0; Mismatches 37; Indels 3; Gaps 1;  
 QY 395 GGCCAGGCGCAGTAGTCTCTGCTGATATCCCAACACTTTGGAGGCCAAGGTGGCGGA 454  
 Db 278 GGCCAGGCGCAGTAGTCTGATATCCCAACACTTTGGAGGCCAAGGTGGAGGAGA 219  
 QY 455 TCACCTGAGGTGAGAGTTTAAGACAGAGCTGGCCACATGATGTAACCTGTCTACT 514  
 Db 218 TCACTGAGGTGAGAGTTTAAGAACAGCTGTGATCAATGATGTAACCTCATCTACT 159



QY	477	GACAGCGCTGGCCAAACATGGGAAAACCCCTGCTCTACTAAAAAAAGCAAAAAATTGGCCAGG	536
Db	207	GACCAAGCTGGCCACATGAGAAAACCCCTTCTCTACTAAAAAACAABAAATTAGCTGGG	266
QY	537	TGTGTGTGCATGCACCTGTAGTCTCCCAACTACTCAGAGGTTGGAGGAGAGAAATTGCTTG	596
Db	267	CATGTGTAGGGGTGCACCTGTAATCCACAGCTACTCGGGAGGCTGAGACAGGMAAAATCGCTTG	326
QY	597	AACTAGAGGGTTGAGGTTGTGCAGTAACCCGAG---ATGTCACTGCACTCCAGCCT-GGCA	652
Db	327	AAACCGGAGGCGGAGTTTGACATGAGCGCGAGATCATCTCAGCTCCAGCTCGGGCG	386
QY	653	ACAGAGCAAGACTCCATTAAGACAAACAAGCCTTGAATTGTGAATTGATGTTTACTT	712
Db	387	ACAGAGCAAGACTCATCTAAAAAAGAAAGAAATGAAAACTGTAGACCAATATCTCTC	446
QY	713	ATCTTCATTTAAGAAATTCATCTT	736
Db	447	ATAACACACGTTGCAAAAATCTCT	470

RESULT	32				
A0769842/c					
LOCUS					
DEFINITION					
A0769842	449 bp	DNA	linear	GSS 28-JUL-1999	
HS_3170-B2.G07.TTC	CIT	Approved	Human	Genomic	Sperm library
saprens_genomic	clone	Plate=3170	Col=14	Row=N	genomic survey

ACCESSION	AQ769842
VERSION	AQ769842.1
KEYWORDS	GSS.

SOURCE ORGANISM	Homo sapiens (human)
	Homo sapiens

ORGANISM	<p><i>Homo sapiens</i>  <i>Eukaryota</i>; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  Homidae; <i>Homo</i>.  1 (bases 1 to 449)</p>
REFERENCE	<p>Mahairs,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  1998</p>
AUTHORS	

**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

PUBMED  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L

401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: [fwallace@u.washington.edu](mailto:fwallace@u.washington.edu)  
Clones may be purchased from Research Genetics ([info@resgen.com](mailto:info@resgen.com))  
BAC and Web Server: <http://www.htsc.washington.edu>  
Placer: 3170 row: N column: 14  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 449.

FEATURES	Location/Qualifiers
source	1..449
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/clone="Plate=3170 Col=14 Row=N"
	/sex="male"
	/clone_1lb="CIT Approved Human Genomic Sperm Library D"
	/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

## ORIGIN

Query Match	10.1%;	Score 202.2;	DB 9;	Length 449;
Best Local Similarity	83.1%;	Pred. No. 3e-17;		
Matches 255;	Conservative 0;	Mismatches 48;	Indels 4;	Gaps 2;
<p>QY 380 AAGAGCTTGAATTGGGCGAGCGAGTAGCTCTGTAATCCCAACATTGGGAG 439</p>				

D <sub>b</sub>	445	AAAAA	CAATC	CATGGG	CGTGGG	CCCTAG	TGGG	CTCAT	GCCCTGT	AATCCAC	ACTTTTG	GGAG	386
Q <sub>y</sub>	440	GCCAAG	TGGGCG	GATTC	ACTGAG	GTCA	GAGGTTA	AGAACA	GCAGCCT	TGGCCA	CATGGTGA	499	
D <sub>b</sub>	385	GCCAAG	TGGG	TGATCAC	CTGAGG	TGAGAG	ATTGAGAG	CAAGC	ACTTGG	CCMACT	GTGTGA	326	
Q <sub>y</sub>	500	AACCTGT	CTCTACT	ATAAAAG	CGAAAAA	TTAG	CCAGTGTG	TGGCAT	GCACTGT	AGTTC	559		
D <sub>b</sub>	325	AACCTGT	CTCTACT	ATAAAA	ATACAAAA	TTAG	CCAGTGTG	TGGTGA	CACTGT	ATAATC	266		
Q <sub>y</sub>	560	CCAACT	ACTCAG	GAGGTG	AGGAGAG	AGATTG	CTTTGAA	CTTAG	GAGGTG	AGGTTC	619		
D <sub>b</sub>	265	CCAGT	ACTCGG	AGGCTG	AGGCCA	GAGATTTG	CTTGAT	CCACGAG	GAGGCA	AGGCTAC	206		
Q <sub>y</sub>	620	TAACCC	GAGA---	TGTCA	CTGCAT	CCAGCCT-	GGCAA	CAGA	GAGCA	AGCTCC	ATAAGAC	675	
D <sub>b</sub>	205	TGAAT	GTGAGAG	CTTGCC	ACTGCA	CTCCAC	TCTGGG	CAACAG	AGCAAG	ACTTC	ATTCAAA	146	
Q <sub>y</sub>	676	AACAAA	682										
D <sub>b</sub>	145	AACAAA	139										

RESULT 33	
AQ822778	
LOCUS	
AQ822778	466 bp
	DNA
	linear
	GSS 26-AUG-1999

DEFINITION	ACCESSION
HS_2205_B2A12.MK.C11 sapiens genomic clone sequence.	A0822778

ACCESSION	AQ822778
VERSION	AQ822778.1
KEYWORDS	GSS

VERSION	KEYWORDS
GS5.	A09422118..1
SOURCE	ORGANISM
Homo sapiens	Homo sapiens (human)
Homo sapiens	
Eukaryota, Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria;	Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo	

**TITLE** Sequence-traced connectors: A sequence approach to mapping and  
**AUTHORS** Mamanides, G.G., Wallace, U.S., Smith, R., Osmundson, S.,  
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D.  
 Hood, L.

sequence region	1	2
scanning the human genome		
Proc Natl Acad Sci U.S.A.	96 (17)	9739-9744 (1999)
TOTPNAT.		

COMMENT  
Contact: Manaias GJ, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com) .  
BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 2205 row: B column: 24  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 466.

FEATURES	Location/Qualifiers
source	1. .466

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate2205 Col=24 Row=B"
/sex="male"
/clone_11b="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: paeloadC11; BAC clones in
E-Coli DH10B"

```

ORIGIN

Query Match	10.1%;	Score 202.2;	DB 9;	Length 466;
Best Local Similarity	80.4%;	Pred. No. 3e-17;		



Best Local Similarity 85.6%; Pred. No. 2,8e-17;  
Matches 249; Conservative 0; Mismatches 38; Indels 4; Gaps 2;

QY 396 GCCAGCGCAGTAGCTCCTGCTGTAATCCCAACACTTTGGAGGCCAAGTGGCGGAT 455  
Db 429 GCTGGCGCAGTAGCTCCTGCTGTAATCCCAACACTTTGGAGGCCAAGTGGCGGAT 370  
QY 456 CACCTGAGGTGAGGATTAAAGACCGCTGGCCAAATGTAAGAAACCTGTCTCTACTA 515  
Db 369 CATCTGACCTCAGAGATTGAGACCGCTGGCCAAATGTAAGAAACCTGTCTCTACTA 310  
QY 516 AAAAGCAAAAATTATGCGCAGTGTGGTCATGCACTGTAGTCCCACTACTCAGAGAG 575  
Db 309 AAAATATCAAAATTATGCGCAGTGTGGTCATGCACTGTAGTCCCACTACTCAGAGAG 250  
QY 576 TTGAGGGAGAGAAATTGCTTTGAACCTAGAGGTGAGGTTGCACTAACCCGAGA--TGT 632  
Db 249 CTGAGGCGAGGAATGCTTTGAACCCAGAGGTTGAGGTTGCACTAACCACTGAGT 190  
QY 633 CACTGCACTCCACCTT-GGCAACAGAGCAACTCCATAAGACACACAAA 682  
Db 189 CACTGCACTCCACCTTGGGCTACAGAGCAAGACTCCGTCTCAAAAAAAA 139

RESULT 36  
BUS17124/c 930 bp mRNA linear EST 12-SRP-2002  
LOCUS AGENCOURT\_10164608 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:6514671  
DEFINITION 5', mRNA sequence.

ACCESSION BUS17124 GI:22824650  
VERSION BUS17124  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 930)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM14090 row: F column: 16  
High quality sequence stop: 430.  
Location/Qualifiers  
1. 930

FEATURES  
source  
1. 930  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6514671"  
/tissue\_type="leiomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_71"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 2.1 kb."

## ORIGIN

Query Match 10.1%; Score 202.2; DB 5; Length 930;

Best Local Similarity 82.7%; Pred. No. 2.3e-17;  
Matches 244; Conservative 0; Mismatches 48; Indels 3; Gaps 1;

QY 391 AATGGCCAGGCGCAGTAGCTCCTGCTGTAATCCCAACACTTTGGAGGCCAAGTGGG 450  
Db 328 ATTGGCCAGGCGTGTGGCTCACTTGTATTCACACACTTGGGAGGCCGAGGTGGG 269

QY 451 CGATCACCTGAGGTGAGGATTAAAGCCAGCTTGGCCAAATGTGTAACCCCTGTCTC 510  
Db 268 CGGATCATCTGAGGTGAGGATTAAAGCCAGCTTGGCCAAATGTGTAACCCCTGTCTG 209

QY 511 TACTTAAAAACGCAAAATTTACCCAGTGTGGTGGCACTGTAAGTCCCAACTACTCA 570  
Db 208 TACTTAAAAATCAAAATTTAGCCAGTGTGGTGGCACTGTAAGTCCCAACTACTCG 149  
QY 571 GGAGGTGAGGAGGAATGCTTTGAACCTAGAGGTGAGGTTGCACTAACCCGAG-- 628  
Db 148 GGGGCTTACACAGAGAAATGCTTTGAACCCAGAGGCGGAGGTTGCACTAGCGGAAT 89  
QY 629 -ATGTCACTGCACTCCAGCTTGGCCAAAGAGCAAGCTCCATTAAGACACACAAA 682  
Db 88 CATGCCATGCACTCCAGCTTGGCCAGAGAGAACTCCATCCAAAAAATA 34

RESULT 37  
BF968880 935 bp mRNA linear EST 22-JAN-2001  
LOCUS BF968880  
DEFINITION 602270854F1 NIH\_MGC\_84 Homo sapiens cDNA clone IMAGE:4358940 5',  
mRNA sequence.

ACCESSION BF968880 GI:12336106  
VERSION BF968880  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 935)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM9997 row: 1 column: 13  
High quality sequence stop: 630.  
Location/Qualifiers  
1. 935

FEATURES  
source  
1. 935  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4358940"  
/tissue\_type="adrenal cortex carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_84"  
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site\_1:  
NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dt  
primed. Average insert size 1.229 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 10.1%; Score 202; DB 2; Length 935;

Best Local Similarity 76.5%; Pred. No. 2.4e-17;  
Matches 274; Conservative 0; Mismatches 80; Indels 4; Gaps 2;

QY 347 AAAATTTAAACCTATACACTTAATTTAAAGAGCTTTGAATGGCCAGCGCAG 406  
Db 73 AATTCGTATATATTTCCTATATGAAGTAAAGTAATCAAAATTCGCGCAGACACGG 132  
QY 407 TAGCTCTGCTGCTTATCCCAACACTTTGGAGGCCAAGTGGCGGATACCTGAGGTC 466  
Db 133 TGGCTCACGCGCTGTGATCCAGACACTTTGGAGGCTGAGGTGATGATGATCACTGAGGTC 192

```

OY 467 AGAGTTTAAGACACAGCTGGCCCAACATGGTGAACCTGTCTCTACTAAAAACGAAA 526
    |||||
DB 193 AGAGTTCAAGACACAGCTGGCCCAACATGGTGAACCTGTCTCTACTAAAAATACAAA 252
OY 527 ATTAGCAGAGTGTGTGGCATGACCTGTAGTCCCACTACAGAGGTTGAGGAGCA 586
    |||||
DB 253 ATTAGCAGAGCTGTGTGGCGCACACCTGTATCCAGCTATTTCAGAGAGTGAGACAGGA 312
OY 587 GAATGCTTGAACCTAGAGAGTGGAGTTGCACTAACCCGAGAT---GTCACTGCACCTCC 643
    |||||
DB 313 GAATGCTTGAACCTAGAGAGTGGAGAGTGGAGTTGCACTGAGATCAAGCCACAGCACTCC 372
OY 644 ACCCT-GGCAGACAGACAGACTCCATAAGACAAAGATTGTAATGTGTA 700
    |||||
DB 373 ACCCTGGCAGACAGACAGACTCCATCTTAAAAAAGAGTCAAAATTCTAGAAA 430

RESULT 38
AL712324 439 bp mRNA linear EST 04-SEP-2003
LOCUS DKFZp686E078.f1.686 (synonym: hicc3) Homo sapiens cDNA clone
DEFINITION DKFZp686E078.5', mRNA sequence.
ACCESSION AL712324
VERSION AL712324.1 GI:19695679
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE 1 (bases 1 to 439)
AUTHORS Mambut,R., Heubner,D., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Mambut,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuberg, Germany
This is the 5' sequence of the clone insert.
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email:s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp686E078) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1..439
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686E078"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hicc3)"
/notes="Vector: pTIP16x2; Site_1: SfiI; Site_2: SfiIB;
cDNA-collection"
ORIGIN
Query Match 10.1%; Score 201.8; DB 1; Length 439;
Best Local Similarity 83.0%; Pred. No. 3.4e-17;
Matches 254; Conservative 0; Mismatches 48; Indels 4; Gaps 2;
OY 381 AGAGCTTTGAATGGGCGCAGGCGCAGTACCTCTGCTTAAATCCCAACATTTGGGAGG 440
    |||||
DB 132 AAAACATGAACAGGCGCAGGCGCAGAACTCAACCTTAATCCCAACATTTGGGAGG 191
OY 441 CCAAGTGGCGGAGTCACTGAGGTGAGAGTTTAAGACACAGCTGGCCCAACATGTGA 500
    |||||
DB 192 CCAAGGACAGGTGCTCACTTGAAGTCAAGAGTTTGAACACAGCTGGCCCAACATGTGA 251
OY 501 ACCGTGTCTACTAAAAACGAAAATTAAGCAGGTGTGTGGCATGACCTGTGTGTC 560

```

```

DB 252 ACCGTGTCTACTAAAAATACAAAATTAAGCTGGGTATGGTGGGACCTGTATCT 311
OY 561 CAACACTACTAGAGAGTTGAGGAGAGAAATTTGTAACCTTGAAGAGGTGAGGTTGCACT 620
    |||||
DB 312 CAGTTACTTTGAGAGGCTGAGGAGAGAAATCCCTTGAACTGGAGGTGAGGTTGCACT 371
OY 621 AACCCGAG---TGTCACTGCACTCAGCT--GGCAAGAGACAGACTCCATAAGACA 676
    |||||
DB 372 GAGCCAGAGTTGTGCCACTGCACTCCAGCTGGGCAACAGATTAAGACTTGTCTCAATA 431
OY 677 AAAAA 682
    |||||
DB 432 AAAAA 437

RESULT 39
AG037584 662 bp DNA linear GSS 01-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-013002.R, genomic survey sequence.
DEFINITION AG037584
ACCESSION AG037584.1 GI:16564457
VERSION AG037584.1 GI:16564457
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE ToTokI,Y., Watanabe,H. and Sakaki,Y.
JOURNAL BAC end sequences of library PTB
REFERENCE 2 (bases 1 to 662)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE ToTokI,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suenhiro-chou,Tsukuba-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpsgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
COMMENT
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
location/Qualifiers
1..662
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-013002.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
ORIGIN
Query Match 10.1%; Score 201.8; DB 10; Length 662;
Best Local Similarity 76.5%; Pred. No. 2.9e-17;
Matches 261; Conservative 0; Mismatches 77; Indels 3; Gaps 1;
OY 321 AGAGTTATTTATCAACTATTATTAATTAATTAATTAATTAATTAATTAATTAATTA 380
    |||||
DB 321 AAAAGCTTTTATAGGCAATTAGAACAGACATTTTCAATTTAGTACATTACATCAATCAG 380
OY 381 AGAGCTTTGAATGGGCGCAGGCGCAGTACCTCTGCTTAAATCCCAACATTTGGGAGG 440
    |||||
DB 381 AAATCTAGAGAGGCGCAGGCGCAGTACCTCTGCTTAAATCCCAACATTTGGGAGG 440

```

OY	441	CCAAAGTGGGGGAGTCACTCTGAGGTCAAGAGATTAAAGCCAGAGCTGGCAACATGAGTAA	500
Db	441	CCAAAGTGGGGGAGTCACTTGAAGTCAAGAGATTGCAAGACCAGAGCTGGCAACATGAGTAA	500
OY	501	ACCCGTCTCTACTAATAAACGCAAAATAATGACAGGTGTGTGGCATGCACTCTTAAGTCC	560
Db	501	ACCCGTCTTTACTAATAATAAATAAATAATGAGTGGGCTTGGTGGGTGTGCTGTAATCC	560
OY	561	CAACTACTCAGAGAGGTGAGGGAGAGAAATTGCTTGAACCTAAGAGAGTGGAGTTGACGT	620
Db	561	CAGCTACTTGGAGAGGCTGAGGACGAGGAATTTGTTGAACCCAGAGGAGAGAGTTGAAGT	620
OY	621	AACCCGAGT---GTCACTGCACACTCAGAGCTGGCAACAGAG	658
Db	621	GAGCAAAAGTGGCCGCAATTGCACCTCAAGCTGTGGGGGACAG	661

RESULT 40	AV729893	678 bp	mRNA	linear	EST 22-NOV-2004
LOCUS	AV729893				
DEFINITION	AV729893	HTE	Homio sapiens	CDNA clone HTEALE09 5',	mRNA sequence.
ACCESSION	AV729893				
VERSION	AV729893.2	GI:55949949			
KEYWORDS	EST.				
SOURCE	Homio sapiens				
ORGANISM	Homio sapiens (human)				

REFERENCE  
AUTHORS  
1 (bases 1 to 678)  
Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,  
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L.,  
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,  
Chen, J., Chen, Z., and Han, Z.  
Homo sapiens cDNA HTE clones  
Unpublished (2000)  
On Oct 17, 2000 this sequence version replaced gi:10839314.

**TITLE** Homo sapiens cDNA HTE clones  
**JOURNAL** Unpublished (2000)  
**COMMENT** On Oct 17, 2000 this sequence version replaced gi:10839314.  
**Contract:** Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES	location/Qualifiers
source	1. .678

```

source
1. 678
   /organism="Homo sapiens"
   /mol_type="RNA"
   /db_xref="taxon:9606"
   /clone="HTAIE09"
   /tissue_type="Hypothalamus"
   /dev_stage="Adult"
   /lab_host="SOLR"
   /clone_lib="HTF"
   /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

```

ORIGIN					
Query Match	10.1%	Score 201.8;	DB 1;	Length 678;	
Best Local Similarity	79.3%	Pred. No. 2.9e-17;			
Matches 264;	Conservative	0;	Mismatches 64;	Indels 5;	Gaps 2.

QY 394 GGGCAGACCGAGTAGCTCCTGCCTGTAAATCCCAACCTTTGGGAGAGCCAAAGTGGGCGG 453

Db 239 GGGCGGGGTGACAGTGGCTCAAGCCTGTAAATCCCAACCTTTGGGAGAGCTAGGACAGGTGG 298

QY 454 ATCACTCGAGGTCAAGAGATTAAACACAGCCTGGCAACATAGGTGAACCTGTCTTAC 513

Db 299 ATCACTCGAGGTCAAGAGATTGAACACAGCCTGGCAGCAGCATGTGAACCTGTCTTAC 358

QY 514 TAAATAAGCAAAAATTAGCCAGGTGTGTGTGCACTGTAAGTATCCCACTAATCTCAGA 573

Db 359 TAAATAATATAAGTTACTCGGGGTGTGGGATCACCCTGTAATCCAGTACTCAGGA 418  
 QY 574 GGTTCAGGAGAGAGAAATTCCTTGAACCTAGAGGTTGGAGCTTACAGTAACCCGAGA---T 630  
 Db 419 GGCTGAGGCAAGAGAAATCGCCCTGTACCCAGAGGTGGAGGTTGAAGTGAAGTGCAGATTTGT 478  
 QY 631 GTCACTGCACTCCAGCC--TGCGACAAGAGCAAGACTCCATATAAGCAACAACAAAGCTTTG 688  
 Db 479 GCTACTGCACCTCCAGCCAGNGTGAACAAAGGCGAGACTCATCTCAAAAAACANAAACAAA 538  
 QY 689 AAATTGTGTAATGAGTTGTACCTATCTTCATT 721  
 Db 539 AAGTTATATAATCAATTTAAGATTAATTTCAAT 571

RESULT 41	1102 bp	EST 05-FEB-2002
BM458632/c	mRNA	linear
LOCUS		
DEFINITION	AGENCOURT 6413734 NIH_MGC_85 Homo sapiens CDNA clone IMAGE:5497695	
	5', mRNA sequence.	

ACCESSION	BM458632
VERSION	BM458632.1
KEYWORDS	GI:18507672
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
JOURNAL	Hominidae; Homo.
COMMENT	1 (bases 1 to 1102)
	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
	National Institutes of Health, Mammalian Gene Collection (MGC)
	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.

Email: [cgapds-1@uall.hawaii.gov](mailto:cgapds-1@uall.hawaii.gov)  
Tissue Procurement: Lou Staub  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MSC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at:  
<http://image.liml.gov>  
Plate: LIML1218 row: d column: 16  
High quality sequence stop: 679.

```

FEATURES
source
location/quadrants
1. .1102
/auratam "Horo canjone"

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5497695"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_85"
/note="Organ: lymph; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC Library."

```

ORIGIN	
Query Match	10.1%; Score 201.8; DB 3; Length 1102;
Best Local Similarity	73.0%; Pred. No. 2.4e-17;
Matches 287; Conservative	0; Mismatches 102; Indels 4; Gaps 2;

[illegible]



Db 345 GCAGATGAGTACCTGGGGTCAAGAGTTCAAGACCACTGGCCCAACATGTTGAACCTT 286  
 506 GTCTACTAAAAACGCAAAATTAAGCCAGGTGGTGGCATGCACTGATGTCAGTCCACT 565  
 285 GTCTCTACTAAAAATACAAAATTAATGCGGACATGAGGCGGTGGCTGTGATCCAGCT 226  
 566 ACTCAGAGAGTTGAGGAGAGAAATGCTTGAACCTAGAGGTGAGTTGCAATACCC 625  
 225 ACTCGGAGGCTGAGGAGAGAAATCACTTGAACCTGGAGGCGGAGTTGCAATAGCC 166  
 626 GAGAT--GTCACCTGCACTCCAGCTT--GGCAACAGAGCAAGACTCCATTAAGCAACAA 661  
 165 GAGATTGTTCTCACTGCACTCCAGCTGGGCGAGCAAGAAAGACTTCATCTCAAAACAAC 106  
 682 AGCTTGAATGTGTAAATGAGTTGACTAT 714  
 Db 105 AACCACAAAATGATGTCACATCAAAACCAT 73

## RESULT 42

LOCUS AM408767 456 bp mRNA linear EST 16-FEB-2000  
 DEFINITION UI-HF-BW0-adw-h-10-0-UI.r1 NIH\_MGC\_38 Homo sapiens cDNA clone  
 IMAGE:3063402 5', mRNA sequence.

ACCESSION AM408767  
 VERSION AM408767.1 GI:6927824

KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 456)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov

REFERENCE  
 AUTHORS NIH-MGC  
 TITLE Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 JOURNAL CDNA Library Preparation: M.B. Soares Lab  
 COMMENT CDNA Library Arrayed by: M.B. Soares Lab  
 DNA Sequencing by: M.B. Soares Lab  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www.bio.lnl.gov/bdrip/image/image.html  
 The following repetitive elements were found in this cDNA sequence:  
 121-410, >ALU

## FEATURES

source

Location/Qualifiers  
 1..456  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3063402"  
 /tissue\_type="lymph"  
 /cell\_type="germinal center B cells"  
 /cell\_line="MGC85"  
 /lab\_host="DH10B (LT1)"  
 /clone\_lib="NIH\_MGC\_38"  
 /note="Vector: pT73-Pac; Site\_1: NotI; Site\_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (2.5-3.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

## ORIGIN

Query Match 10.1%; Score 201.6; DB 1; Length 456;  
 Best Local Similarity 81.6%; Pred. No. 3.6e-17;  
 Matches 258; Conservative 0; Mismatches 54; Indels 4; Gaps 2;

QY 371 AATATTAAGAGCTTTGAATGGCCAGGCGCAGTAGTCTCTGATATCCACA 430  
 Db 97 ACACATTAAGAGTTTAAAGGGGGCCGAGTGGCTACGCTGTATATCCAGCA 156  
 QY 431 CTTTGGAGGCCAAGGTGGCGGATCACTGAGTCAAGAGTTTAAGACCAAGCTGGCCA 490  
 Db 157 CTTTGGAGGCTAGAGCGGGCGGATCACTGAGTCAAGAGTTTCAAGACCAAGCTGGCCA 216  
 QY 491 ACATGGTGAACCTGCTCTACTAATAAAGCAAAATTAAGCAGGTTGGTGGCATGCA 550  
 Db 217 ACTTGGGAAACCTGCTCTACTAATAAATTAAGCAAGTGTGATACAGGCA 276  
 QY 551 CCTGTAGTCCCACTACTCAGAGTTGAGGAGGAGAAATGCTTGAACCTAGAGGTGG 610  
 Db 277 CCGTATTTCCACTCTCTGAGTCTGGGGCAGGAGAAATCGTCAAAACCAAGAGGTGG 336  
 QY 611 AGGTGCACTAACCCAGAT--GTCATGCACTCCAGCTT--GGCAACAGAGCAAGATCTC 666  
 Db 337 AAGTTCAGTAAGCTGATGATCAACACACTGTACTCCAGCTGGGCGCAGAGCAAGACTC 396  
 QY 667 CATTAAGACACAAA 682  
 Db 397 TATCTCAAAAAAAA 412

## RESULT 43

LOCUS A1791156/c 524 bp mRNA linear EST 02-JUL-1999  
 DEFINITION AB53e09.x5 StrataGene lung carcinoma 937218 Homo sapiens cDNA clone  
 IMAGE:844552 3' similar to contains Alu repetitive element; contains element TAR1 repetitive element ;, mRNA sequence.

ACCESSION A1791156  
 VERSION A1791156.1 GI:5338872  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 524)  
 NCI/NIH-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute / National Institute of Dental Research,  
 Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished (1997)

REFERENCE  
 AUTHORS NCI/NIH-CCAP  
 TITLE National Cancer Institute / National Institute of Dental Research,  
 JOURNAL Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 COMMENT Other\_ESTs: ab53e09.y5  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov

REFERENCE  
 AUTHORS NCI/NIH-CCAP  
 TITLE National Cancer Institute / National Institute of Dental Research,  
 JOURNAL Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 COMMENT Other\_ESTs: ab53e09.y5  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov

FEATURES  
 source  
 Location/Qualifiers  
 1..524  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:844552"  
 /tissue\_type="lung carcinoma"  
 /cell\_line="NCI-H69"  
 /dev\_stage="cell line NCI-H69"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="StrataGene lung carcinoma 937218"  
 /note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Small cell carcinoma cell line NCI-H69. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCG 3' -3' adaptor sequence: 5' CTCGAGTTTATTTTATTTT 3'"

FEATURES  
 source  
 Location/Qualifiers  
 1..524  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:844552"  
 /tissue\_type="lung carcinoma"  
 /cell\_line="NCI-H69"  
 /dev\_stage="cell line NCI-H69"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="StrataGene lung carcinoma 937218"  
 /note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Small cell carcinoma cell line NCI-H69. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCG 3' -3' adaptor sequence: 5' CTCGAGTTTATTTTATTTT 3'"

## ORIGIN

Query Match 10.1%; Score 201.6; DB 1; Length 524;  
Best Local Similarity 82.4%; Pred. No. 3.4e-17;  
Matches 244; Conservative 0; Mismatches 49; Indels 3; Gaps 1;

377 TTAAGAGCTTTGAAATGGGCGCAGGCGAGTAGCTCTGCTGTAAATCCCAACACTTTGG 436  
DB TAAATTAACATATTAATGAAGCGGGGGTGTGCTCATGCTGTATATCCAGACACTTTGG 243  
QY GAGGCCAAGGTGGCGGATCACTGAGTCAAGAGTTTAAAGCCAGCTGGCCAAATG 496  
DB GAGGCCAAGGTGGCGGATCACTGAGTCAAGAGTTTAAAGCCAGCTGGCCAAATG 183  
QY TGAAGCCCTGCTCTCTAATAAAGCAAAATTAGCAGGTGTGGGATGACCTGTGA 556  
DB TGAAGCCCTGCTCTCTAATAAATTACAAATTAAGTAGGCAATGTGTGTGCTGTGA 123  
QY GTCCCAACTACTCAGAGGTTTGAAGAGAGAGAAATTCCTGTGAACCTTGAAGTGAAGTTG 616  
DB CTCTTAAGTACTAGAGAGGCTGAGGAGAGAGATGTGTTGAACCCAGAGGCGGAGTTG 63  
QY CAGTAAACCCGAGA---TGTCACTGCACTCCAGCTGGCAACAGACAACTTCAT 669  
DB CAGTGAAGCCGAGATGTGTCACCTGCACTCCAGCTGGCGACAGAGCAAACTTCAT 7

RESULT 44  
BG545785 714 bp mRNA linear EST 04-APR-2001  
LOCUS 602573059F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4701302 5',  
DEFINITION mRNA sequence.

ACCESSION BG545785  
VERSION BG545785.1 GI:13544450  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE 1 (bases 1 to 714)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
<http://image.llnl.gov>  
Plate: LHCW1536 row: e column: 15  
High quality sequence stop: 714.

## FEATURES

source

1..714  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4701302"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH\_MGC\_77"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1:  
SfiI (ggccgctcgagc); Site 2: SfiI (ggccatctggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGCGCATATGAGC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGCGAGCGCGGCGGATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.9  
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 10.1%; Score 201.6; DB 2; Length 714;  
Best Local Similarity 82.4%; Pred. No. 3e-17;  
Matches 244; Conservative 0; Mismatches 49; Indels 3; Gaps 1;

390 AATGCGCGGCGAGTAGCTCTGCTGTATATCCCAACACTTTGGAGGCCAAGTTGG 449  
DB AAGAGGCGTGGCGACGCTGCTCATGCCGTGATCCCGTCACTTTGGAGCGTAGTTGG 419  
QY GCGGATCACTAGGTCAAGGATTTAAAGCAAGCCCTGGCGCAATGTGTAAACCCGTCT 509  
DB GTGGATCACTAGGTCAAGGATTTAAAGCAAGCCCTGGCGCAATGTGTAAACCCCATCT 479  
QY CTACTTAAATACCAAAATTTAGCCAGTGTGTGATCAGCTGTAGTCCCACTACTTC 569  
DB CTACTTAAATACCAAAATTTAGCCAGTGTGTGATCAGCTGTAGTCCCACTACTTC 539  
QY AGAGGTTGAGGAGAGAAATTCCTTGAACCTTGAAGGTTGACGTTGACCCGAGA 629  
DB GAGAGGCTGAGGAGAGAAATTCCTTGAACCCAGAGAGTGTGAGTGTGAGTGTGAGA 599  
QY 630 ---TGTCACTGCACTCCAGCTGGCAAGAGAGAGATCCATAAAGACAAAAA 682  
DB TGTGCACTGCACTCCAGCTGGCGACAGAGAGAGATCCATAAATATAA 655

RESULT 45  
CR936826 3181 bp mRNA linear HTC 23-FEB-2005  
LOCUS CR936826  
DEFINITION Homo sapiens mRNA, cDNA DKFZp686K21248 (from clone DKFZp686K21248).  
ACCESSION CR936826  
VERSION CR936826.1 GI:60219737  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE 1 (bases 1 to 3181)  
AUTHORS Ottenwaelder, B., Obermaier, B., Deutschenbauf, S., Schapp, A.,  
Mewes, H.W., Weill, B., Amd, C., Osanger, A., Fobo, G., Han, M. and  
Wiemann, S.  
COMMENT The German cDNA Consortium  
Submitted (22-FEB-2005) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuberberg, GERMANY  
Direct Submission  
Submitted (22-FEB-2005) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuberberg, GERMANY  
Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: [s.wiemann@dkfz-heidelberg.de](mailto:s.wiemann@dkfz-heidelberg.de).  
sequenced by Medigenomix (Martinsried, Germany) within the cDNA  
sequencing consortium of the German Genome Project.  
This clone (DKFZp686K21248) is available at the RZPD Deutsches  
Reessourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.  
Please contact RZPD for ordering:  
<http://www.rzpd.de/cgi-bin/products/c1.cgi?cloneID=DKFZp686K21248>  
Further information about the clone and the sequencing project is  
available at <http://mips.gsf.de/projects/cdna/>.

## FEATURES

source

1..3181  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp686K21248"  
/tissue\_type="salivary gland"  
/clone\_lib="686 (synonym: hlcc3). Vector psport1\_Sfi; host  
DH10B; sites SfiIa + SfiIb"  
/dev\_stage="adult"

## ORIGIN

Query Match 10.1%; Score 201.6; DB 4; Length 3181;  
Best Local Similarity 85.0%; Pred. No. 1.e-17;  
Matches 238; Conservative 0; Mismatches 39; Indels 3; Gaps 1;

```
QY 374 TTATTAAAGAGCTTTGAAATGGGCCAGGCGCAGTAGCTCTGCTGTATATCCCAACTT 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2853 TCATTTGAAATTTTAGATTAGCCCGGGGCGAGTGGCTCAAGCTGTATATCCAGACTT 2912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 434 TGGGAGGCCAAGGTGGGCGGATCACTTAGGTCAGAGTTTAAGACCAGCTTGGCCACA 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2913 TGGGAGGCCGAGGTGGGTGATCAGCTGGGGTCAAGAGTTCAAGACCAGCTGGGCACA 2972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 494 TGGTAAACCCCTGTCTACTATAAAAGCAAAATTAGCCAGGTGGTGGCATGCACT 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2973 TGGTAAACCCCACTTACTATAAAATACAAAATTAGCCAGGTGGTGGCCACT 3032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 554 GTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAACCTAGAGGTGAGG 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3033 GTATCCCACTTACTCAGAGGCTGAGGACAGAGAAATCGTTGAACCAAGGAGCGGAGG 3092
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 614 TTGCAGTAAACCGAGATGTCAC--TGCCTCCAGCCTGG 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3093 TTGCAGTGAACCGAGATCACACCATTTGCCTCCAGCCTGG 3132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: January 22, 2006, 21:08:48  
Job time : 5429.54 secs

This Page Blank (unpft)

This Page Blank (unpft)

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 01:00:32 ; Search time 6808.48 Seconds  
(without alignments)  
16706.187 Million cell updates/sec

Title: US-09-728-552a-3\_COPY\_39000\_41000

Perfect score: 2001

Sequence: 1 tgcaccacagctctgcctc.....ctaatcagtgacacttca 2001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_hcg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2001	100.0	80622	8	AF222855 Homo sapi
2	1962.6	98.1	112442	8	ALJ55340 Human DNA
3	1962.6	98.1	176432	14	ALJ391648 Homo sapi
4	1961	98.0	69058	8	AF222856 Homo sapi
5	1961	98.0	80202	8	AF222854 Homo sapi
6	1959.4	97.9	41008	6	AX033912 Homo sapi
7	1959.4	97.9	40155	8	AF042484 Homo sapi
8	1918	95.9	40917	6	AX033911 Homo sapi
9	683.6	34.2	71032	14	AC084084 Homo sapi
10	233.6	11.7	122076	8	AC088051 Homo sapi
11	224.8	11.2	140842	14	AC144747 Homo sapi
12	224.8	11.2	173623	14	AC1214872 Homo sapi
13	224.4	11.2	170787	8	AC1214872 Homo sapi
14	224.4	11.2	174363	8	AC135506 Homo sapi
15	224.4	11.2	185146	14	AC104311 Homo sapi
16	224.2	11.2	141521	8	AC073546 Homo sapi
17	224.2	11.2	142838	14	AC026465 Homo sapi
18	224.2	11.2	180807	8	AC022445 Homo sapi
19	223.2	11.2	165467	14	AC159033 Pan trogl
20	222.2	11.1	155533	8	AC079177 Homo sapi
21	222	11.1	106499	8	AC023271 Homo sapi
22	222	11.1	151367	8	AC025750 Homo sapi
23	221.2	11.1	90289	14	AF003164 Homo sapi
24	221.2	11.1	121162	6	AX039602 Sequence
25	221.2	11.1	171612	8	ALJ58013 Human DNA
26	221	11.0	185623	14	AC137058 Pan trogl
27	221	11.0	202317	8	AC020603 Homo sapi
28	220.6	11.0	174965	8	AC036222 Homo sapi
29	220	11.0	154369	8	AC004803 Homo sapi
30	220	11.0	162997	8	AC092035 Homo sapi
31	220	11.0	179146	8	AC067852 Homo sapi
32	219.2	11.0	841	10	BV651164 S217P6186
33	219.2	11.0	152966	8	ALJ59713 Human DNA
34	219.2	11.0	159391	14	AC027113 Homo sapi
35	219.2	11.0	180707	14	AC017038 Homo sapi
36	219	10.9	158135	8	AC025682 Homo sapi
37	219	10.9	248748	14	AC153308 Gorilla g
38	218.8	10.9	102717	8	AC004079 Homo sapi
39	218.8	10.9	133955	8	AC011380 Homo sapi
40	218.8	10.9	178459	14	AC079931 Homo sapi
41	218.8	10.9	179006	8	ALJ36365 Human DNA
42	218.8	10.9	305000	8	HSXDP A
43	218.4	10.9	103681	8	HS568C11 Homo sapi
44	218.4	10.9	181174	8	AC020634 Homo sapi
45	218.4	10.9	187825	14	AC161061 Pan trogl
46	218.4	10.9	191971	8	AC146102 Homo sapi
47	218.2	10.9	102818	14	AP000621 Homo sapi
48	217.6	10.9	81874	8	HS931E15 Human DNA
49	217.6	10.9	120029	14	HS282810 Homo sapi
50	217.6	10.9	154442	14	AC083971 Homo sapi
51	217.6	10.9	157493	8	AC044836 Homo sapi
52	217.6	10.9	190045	14	AF295016 Homo sapi
53	217.6	10.9	201197	14	HS245012 Homo sapi
54	217.2	10.9	164885	14	AC126227 Homo sapi
55	217.2	10.9	170237	8	AC142285 Pan trogl
56	217.2	10.9	193001	14	AC023231 Homo sapi
57	216.6	10.8	135762	14	AC139785 Homo sapi
58	216.6	10.8	139475	8	AC115992 Homo sapi
59	216.6	10.8	162750	8	ALJ55365 Human DNA
60	216.4	10.8	122303	8	AC008681 Homo sapi
61	216.2	10.8	177180	8	HS473B4 Human DNA s
62	216.2	10.8	180064	8	AC025265 Homo sapi
63	215.8	10.8	193104	14	AC090339 Homo sapi
64	215.6	10.8	153170	8	HS1103G7 Human DNA
65	215.6	10.8	159116	14	AC074279 Homo sapi
66	215.4	10.8	141659	8	ALJ39041 Homo sapi
67	215.4	10.8	146376	8	AC009247 Homo sapi
68	215.4	10.8	158608	8	CNS057DJ Homo sapi
69	215.4	10.8	173877	8	AC108671 Human chr
70	215.4	10.8	177273	8	AC074194 Homo sapi
71	215.4	10.8	193803	8	AC022819 Homo sapi
72	215.2	10.8	184435	14	AC141618 Homo sapi
73	215	10.7	176750	8	AC005537 Homo sapi
74	214.8	10.7	137176	8	AC005017 Homo sapi
75	214.8	10.7	165384	14	AC055735 Homo sapi
76	214.8	10.7	170346	8	AC020626 Homo sapi
77	214.8	10.7	187486	14	AC079840 Homo sapi
78	214.6	10.7	152574	8	AC011195 Homo sapi
79	214.6	10.7	153147	8	HS13D10 Human DNA
80	214.6	10.7	157493	14	AC027068 Homo sapi
81	214.6	10.7	168128	14	AC024729 Homo sapi
82	214.6	10.7	170008	8	AC090160 Homo sapi
83	214.6	10.7	172601	14	AC021455 Homo sapi
84	214.6	10.7	191889	14	AC161015 Pan trogl
85	214.6	10.7	195097	8	AC100875 Homo sapi
86	214.4	10.7	170273	8	CNS0509C Homo sapi
87	214.4	10.7	325791	6	AR606191 Sequence
88	214.4	10.7	325791	6	AX234657 Sequence
89	214.2	10.7	180049	8	AC099558 Homo sapi
90	214	10.7	99257	8	ALJ65230 Human chr
91	214	10.7	158467	8	CNS07EF5 Human chr

c 92	214	10.7	161222	14	AC016037	AC016037	Homo sapi
c 93	213.8	10.7	131563	8	HSEB2310	BX322561	Homo sapi
c 94	213.8	10.7	140856	8	AC002430	AC002430	Human BAC
c 95	213.8	10.7	159849	14	AC021454	AC021454	Homo sapi
c 96	213.8	10.7	159670	8	AL390962	AL390962	Human DNA
c 97	213.8	10.7	170842	8	AC007384	AC007384	Homo sapi
c 98	213.8	10.7	180445	8	AC019171	AC019171	Homo sapi
c 99	213.8	10.7	223201	8	HS53110	AL133493	Homo sapi
c 100	213.8	10.7	340000	8	HS21C102	AL163302	Homo sapi
ALIGNMENTS							
RESULT 1							
AF222855							
LOCUS	AF222855	80622	bp	DNA	linear	PRI 17-JUL-2000	
DEFINITION	Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.						
ACCESSION	AF222855						
VERSION	AF222855.1						
KEYWORDS							
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens (human)						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;						
	Hominae; Homo.						
	1 (bases 1 to 80622)						
AUTHORS	Barry,A.E., Bateman,M., Howman,E.V., Cancelli,M.R., Talnton,K.M.,						
	Irvine,D.V., Saffery,R. and Choo,K.H.						
	The 10q25 neocentromere and its inactive progenitor have identical						
	primary nucleotide sequence: further evidence for epigenetic						
TITLE	modification						
	Genome Res. 10 (6), 832-838 (2000)						
	10854414						
	2 (bases 1 to 80622)						
JOURNAL	Barry,A.E.						
REFERENCE	Direct Submisison						
AUTHORS	Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch						
JOURNAL	Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,						
REMARK	Parkville, Melbourne, Victoria 3052, Australia						
	Genomic sequence from human 10q25.2, clone1b-HC						
	Location/Qualifiers						
	1. 80622						
FEATURES	/organism="Homo sapiens"						
SOURCE	/mol_type="genomic DNA"						
	/db_xref="taxon:9606"						
	/chromosome="10"						
	/map="10q25.2"						
repeat_region	/clone="HC"						
	complement(3..176)						
	/rpt_family="Alu"						
	complement(979..3564)						
repeat_region	/r						

/rpt\_family="MIR"  
repeat\_region 65794..65883  
/rpt\_family="MERS"  
repeat\_region 66772..66951  
/rpt\_family="MER20"  
repeat\_region complement(67895..68416)  
/rpt\_family="MER41"  
repeat\_region complement(73192..73470)  
/rpt\_family="Alu"  
repeat\_region 74086..74367  
/rpt\_family="Alu"  
repeat\_region 75916..76183  
/rpt\_family="Alu"  
repeat\_region complement(77771..78059)  
/rpt\_family="Alu"

## ORIGIN

Query Match 100.0%; Score 2001; DB 8; Length 80622;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGATCCACGACCTTG3CCCTCCCAAGTGTGGGATTTACAGCGTGAACCCAGCCGCTG 60  
DB 39027 TGATCCACGACCTTG3CCCTCCCAAGTGTGGGATTTACAGCGTGAACCCAGCCGCTG 39086  
QY 61 GTGCAATGTCTTATTTATTTGAAGAGACAACATGGGCTTAAATCTGTCTTATTTGAC 120  
DB 39087 GTGCAATGTCTTATTTATTTGAAGAGACAACATGGGCTTAAATCTGTCTTATTTGAC 39146  
QY 121 AGACTTGTAGTGAAGTCAAAATCCCAATGTGTCCACTTACTGAACGGCTTAAATGACTTGA 180  
DB 39147 AGACTTGTAGTGAAGTCAAAATCCCAATGTGTCCACTTACTGAACGGCTTAAATGACTTGA 39206  
QY 181 TCTCTCAGCTGTCTTCTGTGATATGTAAGTGTGAATATGATGTGCTTCAAGAGAA 240  
DB 39207 TCTCTCAGCTGTCTTCTGTGATATGTAAGTGTGAATATGATGTGCTTCAAGAGAA 39266  
QY 241 TAAACCTATGAAGAGTGTGAGATAGTGTGATATGAATTAAGATTTCAACAGTGA 300  
DB 39267 TAAACCTATGAAGAGTGTGAGATAGTGTGATATGAATTAAGATTTCAACAGTGA 39326  
QY 301 TAGCTGTATTTGAAGATTTAAGATTAATTTATACAACTATTTAATTAATTTTAAAAAC 360  
DB 39327 TAGCTGTATTTGAAGATTTAAGATTTAATTTATACAACTATTTAATTAATTTTAAAAAC 39386  
QY 361 TAATACCTTAATTAATTTAAGAGCTTTGAATGGGCCAGGCGCAGTAGCTCTGCTGT 420  
DB 39387 TAATACCTTAATTAATTTAAGAGCTTTGAATGGGCCAGGCGCAGTAGCTCTGCTGT 39446  
QY 421 AATCCCAACATTTTGGAGGCGCAAGGTGGCGGATCACTGAGGTCAAGATTTAAGATC 480  
DB 39447 AATCCCAACATTTTGGAGGCGCAAGGTGGCGGATCACTGAGGTCAAGATTTAAGATC 39506  
QY 481 AGCTTGCCCAACATTTGTGAACCCCTGTCTACTATAAAACGCAAAATTTAGCCAGGTGT 540  
DB 39507 AGCTTGCCCAACATTTGTGAACCCCTGTCTACTATAAAACGCAAAATTTAGCCAGGTGT 39566  
QY 541 GTGGCAGTCACTGTAGTCCCACTACTCAGGAGTGTGAAGAGAGAAATGGCTTGAAC 600  
DB 39567 GTGGCAGTCACTGTAGTCCCACTACTCAGGAGTGTGAAGAGAGAAATGGCTTGAAC 39626  
QY 601 TAGGAGTGTGAGTGTGCACTTAACCGGAGATGTCACTGCACTCCAGCTTGCAACAGAGA 660  
DB 39627 TAGGAGTGTGAGTGTGCACTTAACCGGAGATGTCACTGCACTCCAGCTTGCAACAGAGA 39686  
QY 661 AGACTCCATTAAGACAACAAGCTTTGAAATGTGTAAATGAGTGTGTAATCTTTCAT 720  
DB 39687 AGACTCCATTAAGACAACAAGCTTTGAAATGTGTAAATGAGTGTGTAATCTTTCAT 39746  
QY 721 TTAAGAAATTCATCTTTGTTCAATTTATTTTACTTGAATGAGAGCTTCCAGCAATTTT 780  
DB 39747 TTAAGAAATTCATCTTTGTTCAATTTATTTTACTTGAATGAGAGCTTCCAGCAATTTT 39806

QY 781 AATTAAGCCCTCACAGATTTTATGTCACTGTGATGTGAATTAACAATTTATTTGCTAATA 840  
DB 39807 AATTAAGCCCTCACAGATTTTATGTCACTGTGATGTGAATTAACAATTTATTTGCTAATA 39866  
QY 841 TAATATCTTGTCTTCTTTTAAAGAAATGTCTCCCTAGAAAAGTTGTATCCAAACAAT 900  
DB 39867 TAATATCTTGTCTTCTTTTAAAGAAATGTCTCCCTAGAAAAGTTGTATCCAAACAAT 39926  
QY 901 ACACGTACTTTTACACAAATTCAGATGTGATGTGGCAACAGTTGCAATGTTTTCAAAAGAT 960  
DB 39927 ACACGTACTTTTACACAAATTCAGATGTGATGTGGCAACAGTTGCAATGTTTTCAAAAGAT 39986  
QY 961 TTTTCAATTTGAGAAGGGGCCCATTTGGGTATTTAGATTTCAAGAACTGAACCTGTTGT 1020  
DB 39987 TTTTCAATTTGAGAAGGGGCCCATTTGGGTATTTAGATTTCAAGAACTGAACCTGTTGT 40046  
QY 1021 TCTGTTTTCTGCTTCTGAGAGAGAGACATGAATTCAGTTAGCACTTGTGATTT 1080  
DB 40047 TCTGTTTTCTGCTTCTGAGAGAGAGACATGAATTCAGTTAGCACTTGTGATTT 40106  
QY 1081 TCTTTATCTTCAATTTCAATTAAGAAATGCTTCAATGCAAGTGTGTCACTACAT 1140  
DB 40107 TCTTTATCTTCAATTTCAATTAAGAAATGCTTCAATGCAAGTGTGTCACTACAT 40166  
QY 1141 CAAAAGAAAGAGAAACAGTTTCTGTGTTTAAATTTTCAACGGAAAGAAAGGCAACCA 1200  
DB 40167 CAAAAGAAAGAGAAACAGTTTCTGTGTTTAAATTTTCAACGGAAAGAAAGGCAACCA 40226  
QY 1201 TTTTGTCCGCTCTAATTTAGCAGATGATAGTGAAGAGCAGAGTCTTTGAAGGC 1260  
DB 40227 TTTTGTCCGCTCTAATTTAGCAGATGATAGTGAAGAGCAGAGTCTTTGAAGGC 40286  
QY 1261 GTGTATACAGAGTCTTCAATTAATCTCACGAGACTTGCACCTTCAATGCTAAGC 1320  
DB 40287 GTGTATACAGAGTCTTCAATTAATCTCACGAGACTTGCACCTTCAATGCTAAGC 40346  
QY 1321 TGAAGAAATGTCTCAGGAAGATGAACATCTCACAGAGCCCTTAATACTGAAGCAG 1380  
DB 40347 TGAAGAAATGTCTCAGGAAGATGAACATCTCACAGAGCCCTTAATACTGAAGCAG 40406  
QY 1381 TGTATTAAGACACAAGTCAAGAGGCTGAAGAACTAAAGTCTTGAATCTCCACATCTT 1440  
DB 40407 TGTATTAAGACACAAGTCAAGAGGCTGAAGAACTAAAGTCTTGAATCTCCACATCTT 40466  
QY 1441 TCTAGCTCAAGAGGCAAGCTGATTTATTTAGTGAATTTGAATTTTAAATTA 1500  
DB 40467 TCTAGCTCAAGAGGCAAGCTGATTTATTTAGTGAATTTGAATTTTAAATTA 40526  
QY 1501 TTCTAAGTCAGTAACAAAGCCTAATTAATGAATGTGCTGGAAGGTGCTGAATTAAC 1560  
DB 40527 TTCTAAGTCAGTAACAAAGCCTAATTAATGAATGTGCTGGAAGGTGCTGAATTAAC 40586  
QY 1561 TCGATTTTACCAACCCCTCTCTGAGAGAACCATTAATGAATCTGTACATGTTTAC 1620  
DB 40587 TCGATTTTACCAACCCCTCTCTGAGAGAACCATTAATGAATCTGTACATGTTTAC 40646  
QY 1621 TCTACCAAGAACTGTGTTTTCTAATGAGAAACAGAGGCCCAAGATTAATTAAC 1680  
DB 40647 TCTACCAAGAACTGTGTTTTCTAATGAGAAACAGAGGCCCAAGATTAATTAAC 40706  
QY 1681 TTAACCAATACAAATGATAGTGTCTGTGCTTTTATTAAGCACTAAATTTTGAATGC 1740  
DB 40707 TTAACCAATACAAATGATAGTGTGTGCTTTTATTAAGCACTAAATTTTGAATGC 40766  
QY 1741 AATTAATTAATCTGTCCATTAAGAGATTTTCCCTAATGTAATGTAATGTAATGTTCCCT 1800  
DB 40767 AATTAATTAATCTGTCCATTAAGAGATTTTCCCTAATGTAATGTAATGTAATGTTCCCT 40826  
QY 1801 TCAAGGGGCAAGTGTCCGTATCAATAGCTAAATGGAATCTTCTTCAACTACATTAAC 1860  
DB 40827 TCAAGGGGCAAGTGTCCGTATCAATAGCTAAATGGAATCTTCTTCAACTACATTAAC 40886  
QY 1861 CAGAGGGCAGAACTTAATAATGTGTGAATGAATGTAATGTAATGTAATGTAATGTAAT 1920



Db 40887 CAGAGGGCAGAACCTAAATGCTGTGAATGACATTCGTCTTCACATCCAGCAGCACT 40946  
QY 1921 GTTGATTTGACCTCTGACGGGCCACGAGCACTATATGCTCAGATGTTTAACTCA 1980  
Db 40947 GTTGCAATTTGACCTCTGACGGGCCACGAGCACTATATGCTCAGATGTTTAACTCA 41006  
QY 1981 TCTAATTCAGTGAACACTTCA 2001  
Db 41007 TCTAATTCAGTGAACACTTCA 41027

RESULT 2  
AL355340 112442 bp DNA linear PRI 18-MAY-2005  
LOCUS Human DNA sequence from clone RP11-383C6 on chromosome 10 contains  
DEFINITION the TRUB1 gene for Trub pseudouridine (psi) synthase homolog 1 (E.  
coli), a novel gene and a CpG island, complete sequence.

ACCESSION AL355340  
VERSION AL355340.17 GI:21436504  
KEYWORDS HTG; TRUB1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
1 (bases 1 to 112442)  
Direct Submission  
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Jun 17, 2002 this sequence version associated gi:16416169.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr10  
RP11-383C6 is from the library RPC1-11.2 constructed by the group  
of Pieper de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----

FEATURES  
source  
1..112442  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="RZPD:RPC1B753C06383"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-383C6"  
/clone\_1ib="RPC1-11.2"  
2000  
/note="Clone right end. RP11-106M7"  
join(24099..24445,28551..28649,37000..37055,45632..45713,  
56274..56346,58041..58180,60235..60291,61029..63577)  
/gene="TRUB1"

mRNA  
/locus\_tag="RP11-383C6.1-001"  
join(24099..24445,28551..28649,37000..37055,45632..45713,  
56274..56346,58041..58180,60235..60291,61029..63577)  
/gene="TRUB1"  
/locus\_tag="RP11-383C6.1-001"  
/product="Trub pseudouridine (psi) synthase homolog 1 (E.  
coli)"  
/note="match: ESTs: A1208778.1 BQ75335.1  
match: CDNAS: AF448144.1 BC030601.1"  
join(24160..24445,28551..28649,37000..37055,45632..45713,  
56274..56346,58041..58180,60235..60291,61029..61285)  
/gene="TRUB1"  
/locus\_tag="RP11-383C6.1-001"  
/standard\_name="OTTHUMP0000020554"  
/note="match: proteins: Q8MWH5 Q9D331"  
/codon\_start=1  
/product="Trub pseudouridine (psi) synthase homolog 1 (E.  
coli)"  
/protein\_id="CAH73219.1"  
/db\_xref="GI:55653184"  
/db\_xref="Genew:16060"  
/db\_xref="GOA:Q8MWH5"  
/db\_xref="InterPro:IPR002501"  
/db\_xref="InterPro:IPR004510"  
/db\_xref="UniProt/TrEMBL:Q8MWH5"  
/translation="MASEAVALVSSPSSKTTDSPLYLEAGTYAAMAATPSRAAAAV  
AAARTGSEARVSKALATKLISGVAVHPKPTAEALINRLKEKILAEAMPS  
EWTKRKQTLKIGHGTLDSARGLVVGISGCTGMLTSMLSGSRRYAIGLIGATD  
TLDSRGVTEKEKPYDKITQEDIEGLTGNGIMOVPLYSALKDQGRSLTMKRGD  
VVEAKPAPLVVYSISLQKPPFPFTLVEGCGGYYISIVSDIGKEISSLCAVVELT  
RTKQKPTLREHALPBDKTTDIDIOSLEHCSLFPALALKKSPESNEQVLSCEYI  
TLNEKRRDDVTKTC"  
join(24292..24445,28551..28649,37000..37055,45632..45906)  
/gene="TRUB1"  
/locus\_tag="RP11-383C6.1-002"  
join(24292..24445,28551..28649,37000..37055,45632..45906)  
/gene="TRUB1"  
/locus\_tag="RP11-383C6.1-002"  
/product="Trub pseudouridine (psi) synthase homolog 1 (E.  
coli)"  
/note="match: ESTs: A141848.1 A1422325.1 A1640751.1  
BB58241.1"  
61455..61460  
/gene="TRUB1"  
/locus\_tag="RP11-383C6.1-001"  
61479  
/gene="TRUB1"  
/locus\_tag="RP11-383C6.1-001"  
63582  
/locus\_tag="RP11-383C6.1-001"  
63582  
/locus\_tag="RP11-383C6.2-001"  
/locus\_tag="RP11-383C6.2-001"  
join(80563..80691,81349..81444,82381..82499)  
/locus\_tag="RP11-383C6.2-001"  
/product="putative novel transcript"  
/note="match: ESTs: AA74571.1"  
110443  
/note="Clone left end: RP11-338L11"

ORIGIN  
Query Match 98.1%; Score 1962.6; DB 8; Length 112442;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1990; Conservative 0; Mismatches 9; Indels 2; Gaps 2;  
QY 1 TGATCCACGACCTTGCCCTCCCAAGTCTGGGATTACAGCGTGAACCAACCCCTTG 60  
Db 107622 TGATCCACGACCTTGCCCTCCCAAGTCTGGGATTACAGCGTGAACCAACCCCTTG 107681  
QY 61 GTGGAATGCTTTATTTATTTGAAGACAAATGCGCCTTAATCTGCTCTATTTGAC 120  
Db 107682 GTGGAATGCTTTATTTATTTGAAGACAAATGCGCCTTAATCTGCTCTATTTGAC 107741

QY 121 AGACTTGTAGTGAATCCCAATGCTCCACTTAAGAGGCGCTTAATGACTTAG 180  
| | | | |  
Db 107742 AGACTTGTAGTGAATCCCAATGCTCCACTTAAGAGGCGCTTAATGACTTAG 107801  
| | | | |  
QY 181 TCTCTCTGAGCTCTCTTCTGCAATGTAAAGTGAATGAATGAGCTTTCAAGAGAA 240  
| | | | |  
Db 107802 TCTCTCTGAGCTCTCTTCTGCAATGTAAAGTGAATGAATGAGCTTTCAAGAGAA 107860  
| | | | |  
QY 241 TAAACCTATGAAAAGTGTGAAGATGTGTTGATATGAAAATGAAGTTTCAAGAGTAG 300  
| | | | |  
Db 107861 TAAACCTATGAAAAGTGTGAAGATGTGTTGATATGAAAATGAAGTTTCAAGAGTAG 107919  
| | | | |  
QY 301 TACCTGCTATTTGAAGATTTAAGATTTATTTATTAACAATTTTAAATTTTAAAGAA 360  
| | | | |  
Db 107920 TACCTGCTATTTGAAGATTTAAGATTTATTTATTAACAATTTTAAATTTTAAAGAA 107979  
| | | | |  
QY 361 TAAATACATTTAAATTTAAGAGCTTTGAATGAGGCGAGGCGAGTAGCTCCTGCTGT 420  
| | | | |  
Db 107980 TAAATACATTTAAATTTAAGAGCTTTGAATGAGGCGAGGCGAGTAGCTCCTGCTGT 108039  
| | | | |  
QY 421 AATCCCAACATTTTGGAGGCGCAAGTGGCGGATCACTGAGGTCAAGAGTTTAAAGACC 480  
| | | | |  
Db 108040 AATCCCAACATTTTGGAGGCGCAAGTGGCGGATCACTGAGGTCAAGAGTTTAAAGACC 108099  
| | | | |  
QY 481 AGCTGCGCAACATTTGAGGAAACCTGTCTCTACTTAAAGAAAGCAAAATTTAGCGAGGTGTG 540  
| | | | |  
Db 108100 AGCTGCGCAACATTTGAGGAAACCTGTCTCTACTTAAAGAAAGCAAAATTTAGCGAGGTGTG 108159  
| | | | |  
QY 541 GTGGACATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATTTGCTTGAACC 600  
| | | | |  
Db 108160 GTGGACATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATTTGCTTGAACC 108219  
| | | | |  
QY 601 TAGGAGTGAAGGTTGACATGACCCAGATGTCACTGACTCCAGCTGGCAACAGACGA 660  
| | | | |  
Db 108220 TAGGAGTGAAGGTTGACATGACCCAGATGTCACTGACTCCAGCTGGCAACAGACGA 108279  
| | | | |  
QY 661 AGACTCATTAAGAAACAAACCTTTGAAATGTGTAAATGAGTTGATCTTCTGAT 720  
| | | | |  
Db 108280 AGACTCATTAAGAAACAAACCTTTGAAATGTGTAAATGAGTTGATCTTCTGAT 108339  
| | | | |  
QY 721 TTAAGAAATTCATCTTGTTCATTTATTTTACTGTACATGAGAGCTTCAGAAATTTT 780  
| | | | |  
Db 108340 TTAAGAAATTCATCTTGTTCATTTATTTTACTGTACATGAGAGCTTCAGAAATTTT 108399  
| | | | |  
QY 781 AATTAAGCCCTCAAGATTTTATGTCACTGGCTATGTATTAACAATTTATTTGCTAATA 840  
| | | | |  
Db 108400 AATTAAGCCCTCAAGATTTTATGTCACTGGCTATGTATTAACAATTTATTTGCTAATA 108459  
| | | | |  
QY 841 TAAATATCTGTGCTTTTAAAGAAATGTCTCCCTAGAAACGTTTGTACCAACAT 900  
| | | | |  
Db 108460 TAAATATCTGTGCTTTTAAAGAAATGTCTCCCTAGAAACGTTTGTACCAACAT 108519  
| | | | |  
QY 901 ACACTGACTTTACAAAAATCAGATCTGATTTGCAACAGTGCAGATGTTTCAAAAAGAT 960  
| | | | |  
Db 108520 ACACTGACTTTACAAAAATCAGATCTGATTTGCAACAGTGCAGATGTTTCAAAAAGAT 108579  
| | | | |  
QY 961 TTTTCAATTTAGAAAGGGGCCATTTGGTTATTTAGATTTCTAAGAACTGAACTGCTTGT 1020  
| | | | |  
Db 108580 TTTTCAATTTAGAAAGGGGCCATTTGGTTATTTAGATTTCTAAGAACTGAACTGCTTGT 108639  
| | | | |  
QY 1021 TCTGTTTTTCTGCTTCTGGAGAGAGAGAGACATGAAATTCAGATTAGCACTTTGGTATTT 1080  
| | | | |  
Db 108640 TCTGTTTTTCTGCTTCTGGAGAGAGAGAGACATGAAATTCAGATTAGCACTTTGGTATTT 108699  
| | | | |  
QY 1081 TCTTTATCTTCATTTCAATACAGAGATGCTTCATATGCAACAGTGTGTCAGGTACAT 1140  
| | | | |  
Db 108700 TCTTTATCTTCATTTCAATACAGAGATGCTTCATATGCAACAGTGTGTCAGGTACAT 108759  
| | | | |  
QY 1141 CAAAAGAAAGAGAAACAGTTTCTGTGTTTTTAAATTTTCAACCGGAAAGAAAGGCAACCA 1200  
| | | | |  
Db 108760 CAAAAGAAAGAGAAACAGTTTCTGTGTTTTTAAATTTTCAACCGGAAAGAAAGGCAACCA 108819  
| | | | |

QY 1201 TTTTGTCCGCTCTAATTTAGCCAGTGCATGACTTAAGAGAGAGAGAGAGAGTCTTTGAAGGC 1260  
| | | | |  
Db 108820 TTTTGTCCGCTCTAATTTAGCCAGTGCATGACTTAAGAGAGAGAGAGAGAGTCTTTGAAGGC 108879  
| | | | |  
QY 1261 GTGGTAAACAGAGTCTTCAATTAATCTTCCACGAGAGCTTGCACTTCTACTAGTCTTGAAGC 1320  
| | | | |  
Db 108880 GTGGTAAACAGAGTCTTCAATTAATCTTCCACGAGAGCTTGCACTTCTACTAGTCTTGAAGC 108939  
| | | | |  
QY 1321 TGAAGAAATGCTGCAAGAAATGAAACATCTCCACAGACCCCTAATCACTGAAGCAGG 1380  
| | | | |  
Db 108940 TGAAGAAATGCTGCAAGAAATGAAACATCTCCACAGACCCCTAATCACTGAAGCAGG 108999  
| | | | |  
QY 1381 TGTATTAAGCAACAAGTCAAGAGGAGTGAAGAACTAAGCTTCTTGAATCTCCAGCTTCT 1440  
| | | | |  
Db 109000 TGTATTAAGCAACAAGTCAAGAGGAGTGAAGAACTAAGCTTCTTGAATCTCCAGCTTCT 109059  
| | | | |  
QY 1441 TCTAGTCAAGAGAGCCAGCTGATTTATTTAGTTGAAATTTAAGAAATTTTAAATTA 1500  
| | | | |  
Db 109060 TCTAGTCAAGAGAGCCAGCTGATTTATTTAGTTGAAATTTAAGAAATTTTAAATTA 109119  
| | | | |  
QY 1501 TTTCTAAGTCAAGAAACCTTAATTTAAGATAGTGTGCTGGAAGTCTGAATTAAC 1560  
| | | | |  
Db 109120 TTTCTAAGTCAAGAAACCTTAATTTAAGATAGTGTGCTGGAAGTCTGAATTAAC 109179  
| | | | |  
QY 1561 TCGATTTTCAACACCCCTCTTCTGAGAGAGCCATTAATGGAATCTGTACAAATGTTTAC 1620  
| | | | |  
Db 109180 TCGATTTTCAACACCCCTCTTCTGAGAGAGCCATTAATGGAATCTGTACAAATGTTTAC 109239  
| | | | |  
QY 1621 TCTACCAAGAACTTGTGTTTTTCTAATGAGAAACAGAGGCCACAGATTAATTAAC 1680  
| | | | |  
Db 109240 TCTACCAAGAACTTGTGTTTTTCTAATGAGAAACAGAGGCCACAGATTAATTAAC 109299  
| | | | |  
QY 1681 TTAACCAATTAACAAATGACTAGTGTCTGTCTTTTATTAAGCACTAAATTTGATTC 1740  
| | | | |  
Db 109300 TTAACCAATTAACAAATGACTAGTGTCTGTCTTTTATTAAGCACTAAATTTGATTC 109359  
| | | | |  
QY 1741 AATTAATTAATCTGTCCATTAAGAGAGTTCCTCTAATGATCTGATCTTAATCTTCTCCCT 1800  
| | | | |  
Db 109360 AATTAATTAATCTGTCCATTAAGAGAGTTCCTCTAATGATCTGATCTTAATCTTCTCCCT 109419  
| | | | |  
QY 1801 TCAAGGGGCGACGTGCTCCGTACACATAGCTTAATGAGGACTTCTTCAACTACATTAAC 1860  
| | | | |  
Db 109420 TCAAGGGGCGACGTGCTCCGTACACATAGCTTAATGAGGACTTCTTCAACTACATTAAC 109479  
| | | | |  
QY 1861 CAGAGGGCGAAACCTTAATGCTGTGATGACATTTCTGTTCATCTCAGACAGT 1920  
| | | | |  
Db 109480 CAGAGGGCGAAACCTTAATGCTGTGATGACATTTCTGTTCATCTCAGACAGT 109539  
| | | | |  
QY 1921 GTTGCAATTTGAGCTTCTGAGGGGCAACCGAGACCTATCTGTGAGATGTTTAACTCA 1980  
| | | | |  
Db 109540 GTTGCAATTTGAGCTTCTGAGGGGCAACCGAGACCTATCTGTGAGATGTTTAACTCA 109599  
| | | | |  
QY 1981 TCTAATTCAGTGAACCTTCA 2001  
| | | | |  
Db 109600 TCTAATTCAGTGAACCTTCA 109620  
| | | | |

RESULT 3  
AL391648/c 176432 bp DNA linear HTG 18-JUL-2001  
LOCUS Homo sapiens chromosome 10 clone RP11-203A7, 9 unordered pieces.  
DEFINITION  
ACCESSION AL391648.12 GI:14669309  
VERSION  
KEYWORDS HTG, HTGS\_PHASE1, HTGS\_CANCELLED.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
REFERENCE  
1  
AUTHORS Chapman, J.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,





repeat\_region 27173..27449  
/rpt\_family="Alu"  
repeat\_region 27525..27796  
/rpt\_family="Alu"  
repeat\_region 28702..28978  
/rpt\_family="Alu"  
repeat\_region 30668..30736  
/rpt\_family="MLT1"  
repeat\_region 31460..31993  
/rpt\_family="MER1"  
repeat\_region 32130..32248  
/rpt\_family="MLT1"  
repeat\_region 32685..32961  
/rpt\_family="Alu"  
repeat\_region 33113..33403  
/rpt\_family="Alu"  
repeat\_region 34169..34233  
/rpt\_family="MIR"  
repeat\_region 34639  
/rpt\_family="Alu"  
repeat\_region 35993..36081  
/rpt\_family="MIR"  
repeat\_region 38438..38721  
/rpt\_family="Alu"  
repeat\_region 39081..39333  
/rpt\_family="Alu"  
repeat\_region 42817..42877  
/rpt\_family="MIR"  
repeat\_region 44845..45119  
/rpt\_family="Alu"  
repeat\_region 45532..45665  
/rpt\_family="MIR"  
repeat\_region 46257..46490  
/rpt\_family="Alu"  
repeat\_region 48262..48578  
/rpt\_family="Alu"  
repeat\_region 48687..50337  
/rpt\_family="L1"  
repeat\_region 50349..50611  
/rpt\_family="Alu"  
repeat\_region 50670..50831  
/rpt\_family="L1"  
repeat\_region 51085..51374  
/rpt\_family="Alu"  
repeat\_region 52905..53197  
/rpt\_family="Alu"  
repeat\_region 53467  
/rpt\_family="L1"  
repeat\_region 53896..54467  
/rpt\_family="L1"  
repeat\_region 54929..59723  
/rpt\_family="L1"  
repeat\_region 60018..60302  
/rpt\_family="Alu"  
repeat\_region 60338..60470  
/rpt\_family="MIR"  
repeat\_region 61353..61521  
/rpt\_family="MERS"  
repeat\_region 63416  
/rpt\_family="Alu"  
repeat\_region 63724..63889  
/rpt\_family="MERS"  
repeat\_region 64107..64182  
/rpt\_family="MIR"  
repeat\_region 65459..65548  
/rpt\_family="MERS"  
repeat\_region 66439..66618  
/rpt\_family="MER20"  
repeat\_region 67562..68084  
/rpt\_family="MER41"

ORIGIN  
Query Match 98.0%; Score 1961; DB 8; Length 69058;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1989; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

QY 1 TGATCCACGACCTTGGCTTCCCAAAGTCTGGGATTTACAGGCGTGAGCCACGACCTTG 60  
DB 38659 TGATCCACGACCTTGGCTTCCCAAAGTCTGGGATTTACAGGCGTGAGCCACGACCTTG 38728  
QY 61 GTCCGAATGCTTTATTTATTTAAGAGCAACATGGGCTTTAAATCTGCTCTTATTTGAC 120  
DB 38729 GTCCGAATGCTTTATTTATTTAAGAGCAACATGGGCTTTAAATCTGCTCTTATTTGAC 38788  
QY 121 AGACTTGTATGAGTCATATCCCATGCTGCACCTTACGTAAAGGCTTTAATGA 180  
DB 38789 AGACTTGTATGAGTCATATCCCATGCTGCACCTTACGTAAAGGCTTTAATGA 38848  
QY 181 TCTCTCTGAGCTGTCTTTCTGCATATGTAAAGTGAATTAATGATGCTTTCAAGAGGAA 240  
DB 38849 TCTCTCTGAGCTGTCTTTCTGCATATGTAAAGTGAATTAATGATGCTTTCAAGAGGAA 38907  
QY 241 TTAACCTTATGAAATATGTTGAGATAGGTTGATTAATTAAGAACTTCAACAAGTAG 300  
DB 38908 TTAACCTTATGAAATATGTTGAGATAGGTTGATTAATTAAGAACTTCAACAAGTAG 38966  
QY 301 TAGCTGCTATGAAATTTAAGAGTTATTTATTAACAATTTAATTAATTAATTAATTA 360  
DB 38967 TAGCTGCTATGAAATTTAAGAGTTATTTATTAACAATTTAATTAATTAATTAATTA 39026  
QY 361 TTAATCACTTAATTTATTAAGAGCTTTGAAATGGGCGAGCGCTGCTCTGCTGT 420  
DB 39027 TTAATCACTTAATTTATTAAGAGCTTTGAAATGGGCGAGCGCTGCTCTGCTGT 39086  
QY 421 AATCCCAACATTTGGAGGCGCAAGTGGGCGGATCACCTGAGTCAGAGTTTAAGACC 480  
DB 39087 AATCCCAACATTTGGAGGCGCAAGTGGGCGGATCACCTGAGTCAGAGTTTAAGACC 39146  
QY 481 AGCTGGCCCAACATGGTGAACCTGTCTCTAATAAAGCGAAATTTAGCAGGTGTG 540  
DB 39147 AGCTGGCCCAACATGGTGAACCTGTCTCTAATAAAGCGAAATTTAGCAGGTGTG 39206  
QY 541 GTGGCAGTGCCTGTATGTCCTCAACTACAGAGGTTGAGGAGAGAAATTCCTGAACC 600  
DB 39207 GTGGCAGTGCCTGTATGTCCTCAACTACAGAGGTTGAGGAGAGAAATTCCTGAACC 39266  
QY 601 TAGAGAGTGAAGTTCATTAACCCGAGATGCATGCACTCAGCTGCGCAACAGAGA 660  
DB 39267 TAGAGAGTGAAGTTCATTAACCCGAGATGCATGCACTCAGCTGCGCAACAGAGA 39326  
QY 661 AGACTCCATTAAGAGCAACAAAGCTTGAATTTGTAATGATGTAATCTTATCTCAT 720  
DB 39327 AGACTCCATTAAGAGCAACAAAGCTTGAATTTGTAATGATGTAATCTTATCTCAT 39386  
QY 721 TTAAGAAATTCATCTTTGTCAATTTATTTTACTTGACATGAGAGCTTCCAGCAATTTT 780  
DB 39387 TTAAGAAATTCATCTTTGTCAATTTATTTTACTTGACATGAGAGCTTCCAGCAATTTT 39446  
QY 781 AATTAAGCCCTCAACAATTTTATGTCATGCGGATGTAATTAATTAATTTGCTTAA 840  
DB 39447 AATTAAGCCCTCAACAATTTTATGTCATGCGGATGTAATTAATTAATTTGCTTAA 39506  
QY 841 TAATATTCCTGCTCTTTTAAAGATGTCTCCTAGAAACGTTTGTACCAACAT 900  
DB 39507 TAATATTCCTGCTCTTTTAAAGATGTCTCCTAGAAACGTTTGTACCAACAT 39566  
QY 901 ACACTGACTTTACAAATCAATCTGATTTGCAACAGTTGCAATGTTTCAAAAGAT 960  
DB 39567 ACACTGACTTTACAAATCAATCTGATTTGCAACAGTTGCAATGTTTCAAAAGAT 39626  
QY 961 TTTCAATTTGAAGAGGCGCATTTGGGTTATTTAGATCTTAACAACGTAACGCTTTGT 1020  
DB 39627 TTTCAATTTGAAGAGGCGCATTTGGGTTATTTAGATCTTAACAACGTAACGCTTTGT 39686  
QY 1021 TCTGTTTCTGCTCTGAGAGGAGGAGAGATGATTAATCACTTGAACCTTGTATTT 1080  
DB 39687 TCTGTTTCTGCTCTGAGAGGAGGAGAGATGATTAATCACTTGAACCTTGTATTT 39746

OY	1081	TCCTTATCCCTCATTTGCAATACAGAAAGTGGTCAATATGCAAGTGGTGTCAAGTCAACT	1140
Db	39747	TCCTTATCCCTCATTTGCAATACAGAAAGTGGTCAATATGCAAGTGGTGTCAAGTCAACT	39806
OY	1141	CAAAAGAAAGAGAAACAGTTCCTTGCTTTTAAATTTTCAACCGGAAAGAAAGCAACCA	1200
Db	39807	CAAAAGAAAGAGAAACAGTTCCTTGCTTTTAAATTTTCAACCGGAAAGAAAGCAACCA	39866
OY	1201	TTTTGTCCGCTCTAATTTAGCCAGTGCATGACTTTAGAGCAGGACAGATGCTTTGAAGC	1260
Db	39867	TTTTGTCCGCTCTAATTTAGCCAGTGCATGACTTTAGAGAGAGGACAGATGCTTTGAAGC	39926
OY	1261	GGGGTAACAGAGCTTCATTAATCTCACGCAAGACTTGTCACATTCATATGCTTAGGC	1320
Db	39927	GGGGTAACAGAGCTTCATTAATCTCACGCAAGACTTGTCACATTCATATGCTTAGGC	39986
OY	1321	TGAAGAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAATCTGAAGGCAGG	1380
Db	39987	TGAAGAAATGGCTCAGGAAGATGAACAATCTCACAGAGCCCTAATCTGAAGGCAGG	40046
OY	1381	TGTTATTAAGACAAGTCACAGAGGGTGAAGAACTTAACGTTCTTGAAATCTCCACTCTT	1440
Db	40047	TGTTATTAAGACAAGTCATAGAGGGTGAAGAACTTAACGTTCTTGAAATCTCCACTCTT	40106
OY	1441	TCCTACGTCAAGAGCCCAAGCTGATATATTTAGTTGGAATTTAGAAATTTTAAAAATTA	1500
Db	40107	TCCTACGTCAAGAGCCCAAGCTGATATATTTAGTTGGAATTTAGAAATTTTAAAAATTA	40166
OY	1501	TTCTAATAGTCATGAAACAAGCCTAATTTATTAAGATAGTGGCTGGAAAGGTCTAATATAC	1560
Db	40167	TTCTAATAGTCATGAAACAAGCCTAATTTATTAAGATAGTGGCTGGAAAGGTCTAAGTATAC	40226
OY	1561	TCGATTTTACCAACCCCTCTTCTGAGAGAAAGCCATATGGAATCTGTACATGTTCAAC	1620
Db	40227	TCGATTTTACCAACCCCTCTTCTGAGAGAAAGCCAGATGGAATCTGTATGAAATGTTCAAC	40286
OY	1621	TCCTACCAACGAATCTTGTTTTCTAATAGAGAAACAGAGGCCACAGATTTAACTATTC	1680
Db	40287	TCCTACCAACGAATCTTGTTTTCTAATAGAGAAACAGAGGCCACAGTATGTAATCTATTC	40346
OY	1681	TTTAACCAATACAAATAGCTATAGTCTCTGGTCTTTTATTAACACTTAAATTTTATATTC	1740
Db	40347	TTTAACCAAGAACAAATAGCTATAGTCTCTGGTCTTTTATTAACACTTAAATTTTATATTC	40406
OY	1741	AATATATTAATCTGTCCATTTAGGAAGAGTTTCCCAATATGATCTGGTCTTAATCTTCCGT	1800
Db	40407	AATATATTAATCTGTCCAGTATGAGAGAGTTTCCCTAATATGATCTGGTCTTAATCTTCCGT	40466
OY	1801	TCAAGGGGCCAGTGTCCCGTACACATAGCTAATGGGACTTCTTCAACTACCATTAATTC	1860
Db	40467	TCAAGGGGCCAGTGTCCCGTACACATAGCTAATGGGACTTCTTCAACTACCATTAATTC	40526
OY	1861	CAGAGGGCAGAACTTAAATATGCTGTGAATGACATTCGCTGTTCACATCTCAGCAGCACT	1920
Db	40527	CAGAGGGCAGAACTTAAATATGCTGTGAATGACATTCGCTGTTCACATCTCAGCAGCACT	40586
OY	1921	GTTCGCAATTTAGAGCTTCGACAGGGCCACCCGAGACCTAATATCTGTCTCAGATGTTTAACTCA	1980
Db	40587	GTTCGCAATTTAGAGCTTCGACAGGGCCACCCGAGACCTAATATCTGTCTCAGATGTTTAACTCA	40646
OY	1981	TCCTAATTCAGTGAACACTTCA 2001	
Db	40647	TCCTAATTCAGTGAACACTTCA 40667	
RESULT 5			
AF222854			
LOCUS	AF222854	80202 bp	DNA linear pri 17-JUL-2000
DEFINITION	Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.		
ACCESSION	AF222854		
VERSION	AF222854.1	GI:9246844	
KEYWORDS			
SOURCE	Homo sapiens (human)		

ORGANISM	Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominae; Homo.	
REFERENCE	1 (bases 1 to 80202)
AUTHORS	Barry,A.E., Bateman,M., Howman,E.V., Cancellla,M.R., Tainton,K.M., Irvine,D.V., Satterly,R. and Cho,K.H. The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence; further evidence for epigenetic modification
JOURNAL	Genome Res. 10 (6), 832-838 (2000)
PUBMED	10854414
REFERENCE	2 (bases 1 to 80202)
AUTHORS	Barry,A.E.
TITLE	Direct Submmission
JOURNAL	Submitted (11-JAN-2000) Chromosome Research unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville, Melbourne, Victoria 3052, Australia Human genomic sequence from 10q25.2, clone11b=NC, second release location/Qualifiers
REMARK	
FEATURES	
source	1..80202 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="10" /map="10q25.2" /clone="NC" complement(598..3183) /rpt_family="Tigger1" 2587..2715 /rpt_family="Alu" 3283..3553 /rpt_family="Alu" complement(4548..4809) /rpt_family="Alu" complement(5919..6211) /rpt_family="Alu" 6257..6599 /rpt_family="MT12B2" complement(6848..9123) /rpt_family="Alu" complement(10913..11198) /rpt_family="Alu" complement(12062..12313) /rpt_family="Alu" complement(13957..14236) /rpt_family="Alu" 17133..17424 /rpt_family="Alu" complement(19802..20197) /rpt_family="Alu" complement(22664..22956) /rpt_family="Alu" complement(23935..24113) /rpt_family="MIR" 24585..24884 /rpt_family="Alu" complement(25618..25973) /rpt_family="THR1" 26732..26970 /rpt_family="MER2" 27173..27449 /rpt_family="Alu" 27526..27796 /rpt_family="Alu" 28702..28978 /rpt_family="Alu" complement(30668..30736) /rpt_family="MT1" complement(31460..31993) /rpt_family="MER1" complement(32130..33248) /rpt_family="MT1"

repeat\_region 32685..32961  
/rpt\_family="Alu"  
repeat\_region 33113..33403  
/rpt\_family="Alu"  
repeat\_region 34169..34233  
/rpt\_family="MIR"  
repeat\_region 34332..34639  
/rpt\_family="Alu"  
repeat\_region 35993..36081  
/rpt\_family="MIR"  
repeat\_region 38438..38721  
/rpt\_family="Alu"  
repeat\_region 39081..39333  
/rpt\_family="Alu"  
repeat\_region 42817..42877  
/rpt\_family="MIR"  
repeat\_region 44845..45119  
/rpt\_family="Alu"  
repeat\_region 45532..45665  
/rpt\_family="MIR"  
repeat\_region 46257..46490  
/rpt\_family="Alu"  
repeat\_region 48578..48627  
/rpt\_family="Alu"  
repeat\_region 48687..50337  
/rpt\_family="L1"  
repeat\_region 50349..50611  
/rpt\_family="Alu"  
repeat\_region 50670..50831  
/rpt\_family="L1"  
repeat\_region 51085..51374  
/rpt\_family="Alu"  
repeat\_region 52905..53197  
/rpt\_family="Alu"  
repeat\_region 53896..54467  
/rpt\_family="LTR12"  
repeat\_region 59129..59719  
/rpt\_family="LTR12"  
repeat\_region 60018..60302  
/rpt\_family="Alu"  
repeat\_region 60338..60470  
/rpt\_family="MIR"  
repeat\_region 61353..61521  
/rpt\_family="MERS"  
repeat\_region 63135..63416  
/rpt\_family="Alu"  
repeat\_region 63724..63889  
/rpt\_family="MERS"  
repeat\_region 64107..64182  
/rpt\_family="MIR"  
repeat\_region 65459..65548  
/rpt\_family="MERS"  
repeat\_region 66439..66618  
/rpt\_family="MER20"  
repeat\_region 67562..68084  
/rpt\_family="MER41"  
repeat\_region 72861..73139  
/rpt\_family="Alu"  
repeat\_region 73756..74037  
/rpt\_family="Alu"  
repeat\_region 75588..75853  
/rpt\_family="Alu"  
repeat\_region 77441..77729  
/rpt\_family="Alu"

## ORIGIN

Query Match 98.0%; Score 1961; DB 8; Length 80202;  
Best Local Similarity 99.4%; Pred.No. 0;  
Matches 1989; Conservative 0; Mismatches 10; Indels 2; Gaps 2;  
QY 1 TGATCCACCAAGCTGGCTCCCAAGGCTGGAGTTACAGCGGAGCCACGACCTG 60  
DB 38669 TGATCCACCAAGCTGGCTCCCAAGGCTGGAGTTACAGCGGAGCCACGACCTG 38728

QY 61 GTCAATGCTTTATTTATTTGAAGACCAACATGGGCTTAATCTGCTTTATTTGAC 120  
DB 38729 GTCAATGCTTTATTTATTTGAAGACCAACATGGGCTTAATCTGCTTTATTTGAC 38788  
QY 121 AGACTTGAAGAGTCAATCCCAATGCTGCACTTACTGACGCGCTTAAATGACTTAG 180  
DB 38789 AGACTTGAAGAGTCAATCCCAATGCTGCACTTACTGACGCGCTTAAATGACTTAG 38848  
QY 181 TCTCTCAGCTGCTTTCGCAATATGTAAGTGAATTAATGATGAGCTTTCAGAGGAA 240  
DB 38849 TCTCTCAGCTGCTTTCGCAATATGTAAGTGAATTAATGATGAGCTTTCAGAGGAA 38907  
QY 241 TAAACCTAAGAAAGTGTGAGATAGTGTGATATGAATTAAGATTCAACAATGAG 300  
DB 38908 TAAACCTAAGAAAGTGTGAGATAGTGTGATATGAATTAAGATTCAACAATGAG 38966  
QY 301 TAGCTGTATTAAGATTAAAGATTATTTATTAACAATTATTAATTAATTAATTAAC 360  
DB 38967 TAGCTGTATTAAGATTAAAGATTATTTATTAACAATTATTAATTAATTAATTAAC 39026  
QY 361 TAAATCACTTAATTAATTAAGAGCTTGAATGAGGCGAGGCGAGTCTGCTGCTGT 420  
DB 39027 TAAATCACTTAATTAATTAAGAGCTTGAATGAGGCGAGGCGAGTCTGCTGCTGT 39086  
QY 421 AATCCCAACACTTGGAGGCGCAAGTGGGCGGATCACCTGATGATGAGGTTAAAGAC 480  
DB 39087 AATCCCAACACTTGGAGGCGCAAGTGGGCGGATCACCTGATGATGAGGTTAAAGAC 39146  
QY 481 AGCTGGCCACATGCTGTAACCCCTGCTCTATCTAATAACGAAAAATTAAGCAGGTG 540  
DB 39147 AGCTGGCCACATGCTGTAACCCCTGCTCTATCTAATAACGAAAAATTAAGCAGGTG 39206  
QY 541 GTGGAGTGAACCTGATGCTCAACTACTGAGAGGTGAGAGGAAATTCCTGAACC 600  
DB 39207 GTGGAGTGAACCTGATGCTCAACTACTGAGAGGTGAGAGGAAATTCCTGAACC 39266  
QY 601 TAGAGGTGAGAGGTGACATGAACCCGAGATGTCATCTGCACTGACCTGGCAACAGCA 660  
DB 39267 TAGAGGTGAGAGGTGACATGAACCCGAGATGTCATCTGCACTGACCTGGCAACAGCA 39326  
QY 661 AGACTCCATTAAGACAAACAAAGCTTGAATTTGTAATGATGTAATCTGCTCAT 720  
DB 39327 AGACTCCATTAAGACAAACAAAGCTTGAATTTGTAATGATGTAATCTGCTCAT 39386  
QY 721 TTAAGAAATCATCTTGTCTATTTATTTTACTGACATGAGAGCTTCAGCAATTTT 780  
DB 39387 TTAAGAAATCATCTTGTCTATTTATTTTACTGACATGAGAGCTTCAGCAATTTT 39446  
QY 781 AATTAAGCCTCAAGATTTATGTCATGCTATGATTAACAAATTAATTTGCTAATA 840  
DB 39447 AATTAAGCCTCAAGATTTATGTCATGCTATGATTAACAAATTAATTTGCTAATA 39506  
QY 841 TAAATATCTGCTCTTTTAAAGAAATGTCCTCCAGAAAGGTTTGAACAAACAT 900  
DB 39507 TAAATATCTGCTCTTTTAAAGAAATGTCCTCCAGAAAGGTTTGAACAAACAT 39566  
QY 901 ACATGACTTTACACAAATCAAGATCGATTTGGCAACAGTGCAGATGTTTCAAAAGAT 960  
DB 39567 ACATGACTTTACACAAATCAAGATCGATTTGGCAACAGTGCAGATGTTTCAAAAGAT 39626  
QY 961 TTTCAATTGAGAAAGGCGCAATTTGGTATTTAGATTCTAAGAACTGAACCTGCTTGT 1020  
DB 39627 TTTCAATTGAGAAAGGCGCAATTTGGTATTTAGATTCTAAGAACTGAACCTGCTTGT 39686  
QY 1021 TCTGTTTTTCTGAGCTTCTGGAGAGAGAGACATGAATTCATTAAGCACTTGGTATTT 1080  
DB 39687 TCTGTTTTTCTGAGCTTCTGGAGAGAGAGACATGAATTCATTAAGCACTTGGTATTT 39746  
QY 1081 TCTTTATCTTCATTTCAATTCAGAAAGTCTTATATGACAGTGGTGTGAGTCAAT 1140  
DB 39747 TCTTTATCTTCATTTCAATTCAGAAAGTCTTATATGACAGTGGTGTGAGTCAAT 39806



QY 1141 CAAAAGAAAGAGAAAGAGCTTTCTGGTTTAAATTTTCAACCGGAAAGAAAGCAACCA 1200  
DB 39807 CAAAAGAAAGAGAAAGAGCTTTCTGGTTTAAATTTTCAACCGGAAAGAAAGCAACCA 39866  
QY 1201 TTTTGTTCGGCTTAATTTAGCCAGTGCATGATCTTAAGAGCAGGCGAGATGCTTTGAAGGc 1260  
DB 39867 TTTTGTTCGGCTTAATTTAGCCAGTGCATGATCTTAAGAGCAGGCGAGATGCTTTGAAGGc 39926  
QY 1261 GTGGTAACAGAGCTTTCAATTAATCTCAACGAGGACCTTGACCTTACATAGCCCTAGGc 1320  
DB 39927 GTGGTAACAGAGCTTTCAATTAATCTCAACGAGGACCTTGACCTTACATAGCCCTAGGc 39986  
QY 1321 TGAAGAAATGCGTCAGAGAGATGAACAATCTCAAGAGCCCTACTAATCTGAAGCCAGc 1380  
DB 39987 TGAAGAAATGCGTCAGAGAGATGAACAATCTCAAGAGCCCTACTAATCTGAAGCCAGc 40046  
QY 1381 TGTATTAAGCACAAGTCAAGAGGGGTGAGAACTTAACGTTCTTGAATCTCCCACTTCTT 1440  
DB 40047 TGTATTAAGCACAAGTCAAGAGGGGTGAGAACTTAACGTTCTTGAATCTCCCACTTCTT 40106  
QY 1441 TCTAGCTCAGAGAGCCAGCTGATTAATTTAGTGAATTAAGAAATTTTAAATTA 1500  
DB 40107 TCTAGCTCAGAGAGCCAGCTGATTAATTTAGTGAATTAAGAAATTTTAAATTA 40166  
QY 1501 TTTTAAGTCATGAAACAAGCTTAATTAAGATAGTGTGCTGAAGTGTGAATTAAC 1560  
DB 40167 TTTTAAGTCATGAAACAAGCTTAATTAAGATAGTGTGCTGAAGTGTGAATTAAC 40226  
QY 1561 TCGATTTTACCAACCCCTCTTCTGAGAGAAAGCCATTAATGAATCTGTGAATGTTCAc 1620  
DB 40227 TCGATTTTACCAACCCCTCTTCTGAGAGAAAGCCATTAATGAATCTGTGAATGTTCAc 40286  
QY 1621 TCTACCAAGAACTCTTGTCTTCTAATAGAGAAACAGAGCCCAAGATTAATTAAC 1680  
DB 40287 TCTACCAAGAACTCTTGTCTTCTAATAGAGAAACAGAGCCCAAGATTAATTAAC 40346  
QY 1681 TTAACCAATACAAATAGATAGTGTCTGTCTCTTAAATTAAGCACTAAATTTTATGc 1740  
DB 40347 TTAACCAATACAAATAGATAGTGTCTGTCTCTTAAATTAAGCACTAAATTTTATGc 40406  
QY 1741 AATTAATTAATCTGTCCATTAGAGAGATTTCCCTAAATGTAAGTGTCTTAACTTGTCTT 1800  
DB 40407 AATTAATTAATCTGTCCATTAGAGAGATTTCCCTAAATGTAAGTGTCTTAACTTGTCTT 40466  
QY 1801 TGAAGGGGCGCACTGTCTCCCTAACAATAGCTAAATGAGGACTTCTTCAACATTAAC 1860  
DB 40467 TGAAGGGGCGCACTGTCTCCCTAACAATAGCTAAATGAGGACTTCTTCAACATTAAC 40526  
QY 1861 CAGAGGGCAGAACTTAATAGCTGTGATGACATTCGTGCTGTACATCTCAGCAGCAc 1920  
DB 40527 CAGAGGGCAGAACTTAATAGCTGTGATGACATTCGTGCTGTACATCTCAGCAGCAc 40586  
QY 1921 GTTGCAATTTGAGCTTCTGAGGGCCACCCAGAGACTTAATCTGCTCAGATGTTTAACTA 1980  
DB 40587 GTTGCAATTTGAGCTTCTGAGGGCCACCCAGAGACTTAATCTGCTCAGATGTTTAACTA 40646  
QY 1981 TCTAATTCAGTGAACCTTCA 2001  
DB 40647 TCTAATTCAGTGAACCTTCA 40667

RESULT 6  
AX033912 41008 bp DNA linear PAT 21-sbp-2000  
LOCUS DEFINITION Sequence 4 from Patent WO9851790.  
ACCESSION AX033912  
VERSION AX033912.1 GI:10280480  
KEYWORDS  
SOURCE unidentifed  
ORGANISM unidentifed  
REFERENCE 1  
AUTHORS Cancilla,M.R., Choo,K.H. and Du,S.D.

TITLE A novel nucleic acid molecule  
JOURNAL Patent: WO 9851790-A 4.19-NOV-1998;  
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART  
DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)  
FEATURES  
source  
1. 41008  
/organism="unidentifed"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 97.9%; Score 1959.4; DB 6; Length 41008;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1988; Conservative 0; Mismatches 11; Indels 2; Gaps 2;  
QY 1 TGAATCCACAGAGCTTGGCTCCCAAGTGTGAGATTAAGGCGGTGAGCCACAGCCTG 60  
DB 38738 TGAATCCACAGAGCTTGGCTCCCAAGTGTGAGATTAAGGCGGTGAGCCACAGCCTG 38797  
QY 61 GTGCAATGCTTATTAATTTGAAGAGACAACATGGGCTTAAATCTGTCTTATTTGAC 120  
DB 38798 GTGCAATGCTTATTAATTTGAAGAGACAACATGGGCTTAAATCTGTCTTATTTGAC 38857  
QY 121 AGACTTGTAGAGAGTCAATCCCAATGCTGCCACTTACTGAAACGCGCTTAAATGACTTAA 180  
DB 38858 AGACTTGTAGAGAGTCAATCCCAATGCTGCCACTTACTGAAACGCGCTTAAATGACTTAA 38917  
QY 181 TCTCTCAGCTGTCTTCTGCAATATGTAAGGTGGAATATGATGCGCTTCAAGAGAA 240  
DB 38918 TCTCTCAGCTGTCTTCTGCAATATGTAAGGTGGAATATGATGCGCTTCAAGAGAA 38976  
QY 241 TAAACCTAAGAAAGTGTGAGATGCTGATGATGATGATGATGATGATGATGATGATG 300  
DB 38977 TAAACCTAAGAAAGTGTGAGATGCTGATGATGATGATGATGATGATGATGATGATG 39035  
QY 301 TAGCTGCTATTAAGATTTAAGATTAATTAATTAACAATTAATTAATTAATTAATTA 360  
DB 39036 TAGCTGCTATTAAGATTTAAGATTTAATTAATTAACAATTAATTAATTAATTAATTA 39095  
QY 361 TAATACCTTAATTAATTAAGAGCTTTGAAATGGCGAGCGGAGTACTCTGCTGT 420  
DB 39096 TAATACCTTAATTAATTAAGAGCTTTGAAATGGCGAGCGGAGTACTCTGCTGT 39155  
QY 421 AATCCCAACATTTGGAGGCGCAAGTGGGCGGATCACTGAGGTGAGAGATTAAAGC 480  
DB 39156 AATCCCAACATTTGGAGGCGCAAGTGGGCGGATCACTGAGGTGAGAGATTAAAGC 39215  
QY 481 AGCTGGCCAAATGGTGAACCCCTGTCTCTACTTAAACGCAAAATTAAGCAGGTGTG 540  
DB 39216 AGCTGGCCAAATGGTGAACCCCTGTCTCTACTTAAACGCAAAATTAAGCAGGTGTG 39275  
QY 541 GTGGCATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGGAGAAATTTGTTGAAC 600  
DB 39276 GTGGCATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGGAGAAATTTGTTGAAC 39335  
QY 601 TGAAGAGTGGAGGTTGCACTAACCCGAGATGTCATCTCAGCTCAGCTTGGCAACAGAG 660  
DB 39336 TGAAGAGTGGAGGTTGCACTAACCCGAGATGTCATCTCAGCTCAGCTTGGCAACAGAG 39395  
QY 661 AGACTCCATTAAGAGACAACAAAGCTTTGAAATTTGTAATGATGTTGTAATCTTCAAT 720  
DB 39396 AGACTCCATTAAGAGACAACAAAGCTTTGAAATTTGTAATGATGTTGTAATCTTCAAT 39455  
QY 721 TTAAGAAATTCATCTTGTCTCAATTAATTTTAATCTTGAACATGAGAGCTTCCAGCAATTTT 780  
DB 39456 TTAAGAAATTCATCTTGTCTCAATTAATTTTAATCTTGAACATGAGAGCTTCCAGCAATTTT 39515  
QY 781 AATTAAGCCCTACAGATTTTATGTCAGCTGCTATGTAATTAACAATTAATTTGCTAATA 840  
DB 39516 AATTAAGCCCTACAGATTTTATGTCAGCTGCTATGTAATTAACAATTAATTTGCTAATA 39575  
QY 841 TAATATTCCTGCTCTTTTAAAGAAATGTCCTCCCTAGAAACGGTTGTACCAACAT 900

Db 39576 TAAATCTGCTCTTTTAAAGAAATGTCCTCAAGAAAGCTTTGTAACAAACAT 39635  
 Oy 901 AACTGACTTTTACAAAAATGATCTGATTTGGCAACGTTGCAGATGTTTCAAAAGT 960  
 Db 39636 AACTGACTTTTACAAAAATGATCTGATTTGGCAACGTTGCAGATGTTTCAAAAGT 39695  
 Oy 961 TTTGATTTGAGAGGGGGCCCTTTGGGTTAATTTAGATTTAAGAACTGCTTTGT 1020  
 Db 39696 TTTGATTTGAGAGGGGGCCCTTTGGGTTAATTTAGATTTAAGAACTGCTTTGT 39755  
 Oy 1021 TCTGTTTTCTGCTCTTGGGAGAGAGAGACATGAATTCAGTTAGACCTTTGATTT 1080  
 Db 39756 TCTGTTTTCTGCTCTTGGGAGAGAGAGACATGAATTCAGTTAGACCTTTGATTT 39815  
 Oy 1081 TCTTTATCTCTTCAATTCATACAGAAAGATGCTTCAATTCAGATGCTGCTGCTACAT 1140  
 Db 39816 TCTTTATCTCTTCAATTCATACAGAAAGATGCTTCAATTCAGATGCTGCTGCTACAT 39875  
 Oy 1141 CAAAAAGAAAGAAACAGTTTCTGGGTTTAAATTTTCAACCGSAAAGAAAGCACCA 1200  
 Db 39876 CAAAAAGAAAGAAACAGTTTCTGGGTTTAAATTTTCAACCGSAAAGAAAGCACCA 39935  
 Oy 1201 TTTTGTTCGCTCTAATTAAGCAGTGCATGACTTAGAGACAGCAGATGCTTTGAAGC 1260  
 Db 39936 TTTTGTTCGCTCTAATTAAGCAGTGCATGACTTAGAGAGCAGATGCTTTGAAGC 39995  
 Oy 1261 GTGGTACACAGGCTTTCATTAATCTCCACGCGACCTTGCACTTCTACATAGCCATAGC 1320  
 Db 39996 GTGGTACACAGGCTTTCATTAATCTCCACGCGACCTTGCACTTCTACATAGCCATAGC 40055  
 Oy 1321 TGAAGAAATGGCTGAGAAAGATGAAACAATCTCAGAGGCCCTTAACCTGAAGCCAG 1380  
 Db 40056 TGAAGAAATGGCTGAGAAAGATGAAACAATCTCAGAGGCCCTTAACCTGAAGCCAG 40115  
 Oy 1381 TGTATTAAGCAACAAGTCAAGAGGGTGAGAACTAACTCTTGAATCTCCACTTCTT 1440  
 Db 40116 TGTATTAAGCAACAAGTCAAGAGGGTGAGAACTAACTCTTGAATCTCCACTTCTT 40175  
 Oy 1441 TCTACGTCAGAGAGCCAGCTGATTTTATTTAGTTGAATTTTGAATTTTAAAAATTA 1500  
 Db 40176 TCTACGTCAGAGAGCCAGCTGATTTTATTTAGTTGAATTTTGAATTTTAAAAATTA 40235  
 Oy 1501 TTTCTAAGTCATGAAACAGCCCTAATTAAGATGATGCTGTAAGAGTCTGAAATTAAC 1560  
 Db 40236 TTTCTAAGTCATGAAACAGCCCTAATTAAGATGATGCTGTAAGAGTCTGAAATTAAC 40295  
 Oy 1561 TCGATTTTACCAACCCCTCTCTGAGAGAGCCATTAATGAAATCTGTACAAATGTTTAC 1620  
 Db 40296 TCGATTTTACCAACCCCTCTCTGAGAGAGCCATTAATGAAATCTGTACAAATGTTTAC 40355  
 Oy 1621 TCTACCAAGCACTCTGTTTTTTCTAATGAGAAACAGAGCCCAAGTATTAACCTATC 1680  
 Db 40356 TCTACCAAGCACTCTGTTTTTTCTAATGAGAAACAGAGCCCAAGTATTAACCTATC 40415  
 Oy 1681 TTTAACCAATCAAAATGACTAGTGTCTGAGCTCTTATTAAGCACTAAATTTTGAATC 1740  
 Db 40416 TTTAACCAATCAAAATGACTAGTGTCTGAGCTCTTATTAAGCACTAAATTTTGAATC 40475  
 Oy 1741 AATTAATTAATCTGCTCAATTAAGAGATTTTCCCTAATGTAAGTGTCTTAACTTGTCCCT 1800  
 Db 40476 AATTAATTAATCTGCTCAATTAAGAGATTTTCCCTAATGTAAGTGTCTTAACTTGTCCCT 40535  
 Oy 1801 TCAAGGGGCAAGTGTCCCGTACACATAGCTTAATGGAATCTTCTTCACTACATTAAC 1860  
 Db 40536 TCAAGGGGCAAGTGTCCCGTACACATAGCTTAATGGAATCTTCTTCACTACATTAAC 40595  
 Oy 1861 CAGAGGGCAGAACTTAATAGCTGTGAATGACATTCGCTGTGACATCTCAGAGAGCT 1920  
 Db 40596 CAGAGGGCAGAACTTAATAGCTGTGAATGACATTCGCTGTGACATCTCAGAGAGCT 40655  
 Oy 1921 GTTGCATTTGAGCTTCTGACAGGCGACCCAGAGACTTATATCTGCTCAGATGTTTAACTCA 1980  
 Db 40656 GTTGCATTTGAGCTTCTGACAGGCGACCCAGAGACTTATATCTGCTCAGATGTTTAACTCA 40715

Oy 1981 TCTAATTCAGTGAACACTTCA 2001  
 Db 40716 TCTAATTCAGTGAACACTTCA 40736  
 RESULT 7  
 AF042484  
 LOCUS  
 DEFINITION Homo sapiens sequence of the 80 kb core centromere protein binding region of the mar del (10) neocentromere corresponding to human chromosome 10q25.2, complete sequence.  
 ACCESSION AF042484  
 VERSION AF042484  
 KEYWORDS AF042484.1 GI:4205782  
 SOURCE HTG.  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 80155)  
 AUTHORS Barry,A.E., Howman,E.V., Cancelli,M.R., Saffery,R. and Choo,K.H.  
 TITLE Sequence analysis of an 80 kb human neocentromere  
 JOURNAL Hum. Mol. Genet. 8 (2), 217-227 (1999)  
 PUBMED 9931329  
 REFERENCE 2 (bases 1 to 80155)  
 AUTHORS Barry,A.E., Howman,E.V., Cancelli,M.R., Saffery,R. and Choo,A.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville, Melbourne, Victoria 3052, Australia  
 FEATURES  
 source  
 1..80155  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /map="10q25.2"  
 /note="Sequence information obtained from BAC clones prepared from cell line BE2C1-18-5F"  
 complement(join(149..525,3628..4545,4849..5053))  
 /rpt\_family="LIMC"  
 /rpt\_type=dispersed  
 join(560..597,3184..3281)  
 /rpt\_family="MER58"  
 /rpt\_type=dispersed  
 complement(join(598..2553,2716..3183))  
 /rpt\_family="TIGER1"  
 /rpt\_type=dispersed  
 2558..2581  
 /standard\_name="T1"  
 /note="10% homology"  
 /rpt\_type=direct  
 /rpt\_unit="(cagcacagcgg) 2"  
 2570..2715  
 /rpt\_family="AluSg/x"  
 /rpt\_type=dispersed  
 3021..3050  
 /standard\_name="T2"  
 /note="83.3% homology"  
 /rpt\_type=direct  
 /rpt\_unit="(taacaagtcg) 3"  
 3283..3558  
 /rpt\_family="AluSx"  
 /rpt\_type=dispersed  
 4173..4211  
 /standard\_name="T3"  
 /note="82.1% homology"  
 /rpt\_type=direct  
 /rpt\_unit="(aaaaataatct) 3"  
 complement(4546..4648)  
 /rpt\_family="AluSx"  
 /rpt\_type=dispersed

```
repeat_region complement(5909, .6208)
/rpt_family="AluSg"
/rpt_type=dispersed
join(6232, .6573, 6638, .6688)
/rpt_family="MLT2CB"
/rpt_type=dispersed
6598, .6645
/standard_name="AC24"
/note="100% homology"
/rpt_type=tandem
/rpt_unit="ac"
6598, .6637
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="ca"
7091, .7134
/rpt_family="MIR"
/rpt_type=dispersed
7808, .7911
/rpt_family="L2"
/rpt_type=dispersed
8349, .8524
/rpt_family="MIR"
/rpt_type=dispersed
complement(8839, .9124)
/rpt_family="AluJb"
/rpt_type=dispersed
9180, .9280
/rpt_family="MIR"
/rpt_type=dispersed
10070, .10170
/note="low complexity"
/rpt_family="purine-rich"
10097, .10136
/standard_name="GA1"
/note="82.5% homology"
/rpt_type=direct
/rpt_unit="GaaGaaGaaG" 4"
10175, .10339
/rpt_family="FRAM"
/rpt_type=dispersed
complement(10899, .11197)
/rpt_family="AluSx"
/rpt_type=dispersed
join(11627, .11757, 12434, .12512, 13443, .13562)
/note=">98% identity"
/db_xref="dbEST:AA74571"
11996, .12028
/standard_name="AT4"
/note="100% homology"
/rpt_type=tandem
/rpt_unit="(attc) 8"
complement(12010, .12312)
/rpt_family="AluJb"
/rpt_type=dispersed
12327, .12433
/rpt_family="L2"
/rpt_type=dispersed
complement(13946, .14235)
/rpt_family="AluSg"
/rpt_type=dispersed
15130, .15159
/note="low complexity"
/rpt_family="AT-rich"
15164, .15753
/standard_name="AT28"
/note="82.1% homology"
/rpt_type=direct
/rpt_unit="(atcgtatcgtcatatagacataaact) 21"
join(15170, .15289, 15410, .15612)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="ca"
```

```
repeat_region 15617, .15731
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="cata"
16004, .16100
/rpt_family="L2"
/rpt_type=dispersed
complement(16128, .16364)
/rpt_family="L2"
/rpt_type=dispersed
complement(16407, .16592)
/rpt_family="L2"
/rpt_type=dispersed
complement(16557, .16889)
/rpt_family="MIR"
/rpt_type=dispersed
16914, .16936
/note="low complexity"
/rpt_family="AT-rich"
17102, .17386
/rpt_family="AluSx"
/rpt_type=dispersed
17406, .17564
/standard_name="GA2"
/note="81.8% homology"
/rpt_type=direct
17416, .17543
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="GaaA"
17647, .17702
/rpt_family="MIR"
/rpt_type=dispersed
complement(19863, .20161)
/rpt_family="AluSx"
/rpt_type=dispersed
complement(20842, .20913)
/rpt_family="L2"
/rpt_type=dispersed
complement(21366, .21414)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="GaaA"
21375, .21639
/standard_name="CT1"
/note="75% homology"
/rpt_type=dispersed
complement(21415, .21536)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="TgGaa"
complement(21524, .21646)
/note="low complexity"
/rpt_family="purine-rich"
22405, .22646
/rpt_family="L1M4"
/rpt_type=dispersed
22560, .22604
/standard_name="T4"
/note="80% homology"
/rpt_type=direct
/rpt_unit="(aatcattacataact) 3"
```

Query Match 97.9%; Score 1959.4; DB 8; Length 80155;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1988; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```
QY 1 TGAATCACAGCGCTTGAGCTCCCAAGTGTGGATTACAGCGCTGAGCCACGACCTTG 60
DB 38647 TGAATCACAGCGCTTGAGCTCCCAAGTGTGGATTACAGCGCTGAGCCACGACCTTG 38706
QY 61 GTGCAATGCTTATATTTGAAGACACAATGGGCTTAATCTGTCTTATTTGAC 120
|||||
```

Dh 38707 GTGCAATGTTTTTATTAATTGGAAGACAACAATGGGCCCTTAAATCTGTCATATTGAC 38766  
Qy 121 AGACTTGAATGGAGCAATCCCAATGCGCACTTACTGAACGGGCTTAAATGACTTAC 180  
Dh 38767 AGACTTGAATGGAGCAATCCCAATGCGCACTTACTGAACGGGCTTAAATGACTTAC 38826  
Qy 181 TCTCTGAGCTGTCTTCTGCAATATGTAAGGTGGAATATGATGAGCTTTCAAGAGAA 240  
Dh 38827 TCTCTGAGCTGTCTTCTGCAATATGTAAGGTGGAATATGATGAGCTTTCAAGAGAA 38885  
Qy 241 TAAACCTATGAAAAGTGTGAGATATGTTGATATGTAATATGAGATTTCAACATG 300  
Dh 38886 TAAACCTATGAAAAGTGTGAGATATGTTGATATGTAATATGAGATTTCAACATG 38944  
Qy 301 TAGCGCTATGTAAGATTTAAGATTTATTAACAATTAATTAATTAATTAATTAATTA 360  
Dh 38945 TAGCGCTATGTAAGATTTAAGATTTATTAACAATTAATTAATTAATTAATTAATTA 39004  
Qy 361 TAAATCACTTAATTAATTAAGAGCTTTGAAATGGGCGAGGCGAGTACCTGCTGT 420  
Dh 39005 TAAATCACTTAATTAATTAAGAGCTTTGAAATGGGCGAGGCGAGTACCTGCTGT 39064  
Qy 421 AATCCCAACATTTTGGAGGCGCAAGGTGCGGATCACTGAGGTCAAGATTTAAGACC 480  
Dh 39065 AATCCCAACATTTTGGAGGCGCAAGGTGCGGATCACTGAGGTCAAGATTTAAGACC 39124  
Qy 481 ACCCTGGCAACATGGTGAAACCCCTGCTCTACTAATAAAGCAAAATTAAGCCAGGTG 540  
Dh 39125 ACCCTGGCAACATGGTGAAACCCCTGCTCTACTAATAAAGCAAAATTAAGCCAGGTG 39184  
Qy 541 GTGGACATGACCTGTAGTCCCACTACTGAGAGTTGAGGAGAGAAATGCTTGAAC 600  
Dh 39185 GTGGACATGACCTGTAGTCCCACTACTGAGAGTTGAGGAGAGAAATGCTTGAAC 39244  
Qy 601 TAGAGGTGAGAGTTGCACTTAACCCGAGATGCACTGCACTCCAGCTGGCAACAAG 660  
Dh 39245 TAGAGGTGAGAGTTGCACTTAACCCGAGATGCACTGCACTCCAGCTGGCAACAAG 39304  
Qy 661 AGACTCCATTAAGACAACAAAGCTTGAATGTGTGAATGAGTGTACCTATCTTCAT 720  
Dh 39305 AGACTCCATTAAGACAACAAAGCTTGAATGTGTGAATGAGTGTACCTATCTTCAT 39364  
Qy 721 TTAAGAAATTCATCTTGTGCTTATTTTACTTTCAGATGAGAGCTTCCAGCAATTTT 780  
Dh 39365 TTAAGAAATTCATCTTGTGCTTATTTTACTTTCAGATGAGAGCTTCCAGCAATTTT 39424  
Qy 781 AATTAAGCCCTCAAGATTTATGCACTGGCTATGTGATTAACAATTAATTTGCTAAA 840  
Dh 39425 AATTAAGCCCTCAAGATTTATGCACTGGCTATGTGATTAACAATTAATTTGCTAAA 39484  
Qy 841 TAAATATTTCTGCTTCTTTTAAAGAAATGTCTCCCTAGAAAGGTTTGAACAACAT 900  
Dh 39485 TAAATATTTCTGCTTCTTTTAAAGAAATGTCTCCCTAGAAAGGTTTGAACAACAT 39544  
Qy 901 ACATGACTTACACAAATCAGATCTGATTTGGCAACAGTTCAGATTTTCAAAAGAT 960  
Dh 39545 ACATGACTTACACAAATCAGATCTGATTTGGCAACAGTTCAGATTTTCAAAAGAT 39604  
Qy 961 TTTCAATTGAGAAAGGGCCCATTTGGGTATTTAGATCTTAAGAACTGAACTGCTTGT 1020  
Dh 39605 TTTCAATTGAGAAAGGGCCCATTTGGGTATTTAGATCTTAAGAACTGAACTGCTTGT 39664  
Qy 1021 TCTGTTTTTCTGGCTTGGGAGAGAGAGACATGAATTCAGTTAGCACTTGTATTT 1080  
Dh 39665 TCTGTTTTTCTGGCTTGGGAGAGAGAGACATGAATTCAGTTAGCACTTGTATTT 39724  
Qy 1081 TCTTTATCTTCAATTAACAGAAAGATGCTCATATGCAAGTGTGCTGAGTCAAT 1140  
Dh 39725 TCTTTATCTTCAATTAACAGAAAGATGCTCATATGCAAGTGTGCTGAGTCAAT 39784  
Qy 1141 CAAAAGAAAGAAACAGTTTCTTGTGTTTAAATTTTCAACCGGAAAGAAAGGCAACCA 1200  
Dh 39785 CAAAAGAAAGAAACAGTTTCTTGTGTTTAAATTTTCAACCGGAAAGAAAGGCAACCA 39844

Qy 1201 TTTTGTTCGCTCTAATTAAGCCATGATGACTTGAAGACAGCAGATGCTTGAAGGC 1260  
Dh 39845 TTTTGTTCGCTCTAATTAAGCCATGATGACTTGAAGACAGCAGATGCTTGAAGGC 39904  
Qy 1261 GTGTGAACACAGGTCTTCAATTAATCTCCACGAGAGACTTGCATCTTCAATGCTTAAGC 1320  
Dh 39905 GTGTGAACACAGGTCTTCAATTAATCTCCACGAGAGACTTGCATCTTCAATGCTTAAGC 39964  
Qy 1321 TGAAGAAATGAGCTCAGGAAGATGAACAATCTCAGAGACCCTAATCTAAGACAGG 1380  
Dh 39965 TGAAGAAATGAGCTCAGGAAGATGAACAATCTCAGAGACCCTAATCTAAGACAGG 40024  
Qy 1381 TGTATTAAGCAACAGTCAAGAGGAGTGAAGAACTAACGTTCTTGAATCTCCACTTCT 1440  
Dh 40025 TGTATTAAGCAACAGTCAAGAGGAGTGAAGAACTAACGTTCTTGAATCTCCACTTCT 40084  
Qy 1441 TCTAGCTCAGAAAGACCAAGCTGATTAATTTAGTTGGAATTTAAGAAATTTTAAATTA 1500  
Dh 40085 TCTAGCTCAGAAAGACCAAGCTGATTAATTTAGTTGGAATTTAAGAAATTTTAAATTA 40144  
Qy 1501 TTTCTAAGCATGAACAAGCCTAATTAATTAAGATAGTGTGTAAGGTGCTGAATAAC 1560  
Dh 40145 TTTCTAAGCATGAACAAGCCTAATTAATTAAGATAGTGTGTAAGGTGCTGAATAAC 40204  
Qy 1561 TCGATTTTACCAACCCCTCTTCTGAGAGAACCATTAATGAATCTGTACATGTTTAC 1620  
Dh 40205 TCGATTTTACCAACCCCTCTTCTGAGAGAACCATTAATGAATCTGTGTACATGTTTAC 40264  
Qy 1621 TCTACCAAGCAACTGTTGTTTTTTCTTAATGAAGAAACAGAGGCCACAGATTAACATAC 1680  
Dh 40265 TCTACCAAGCAACTGTTGTTTTTTCTTAATGAAGAAACAGAGGCCACAGATTAACATAC 40324  
Qy 1681 TTAACCAATACAAAATGACTAGTGTCTGCTCTTATTAAGCACTAATAATTTGATCC 1740  
Dh 40325 TTAACCAATACAAAATGACTAGTGTCTGCTCTTATTAAGCACTAATAATTTGATCC 40384  
Qy 1741 AATTAATTAATCTGTCCATTAAGAGAGTTTCCCTAATGTACTGTTCTAATCTTTCCT 1800  
Dh 40385 AATTAATTAATCTGTCCATTAAGAGAGTTTCCCTAATGTACTGTTCTAATCTTTCCT 40444  
Qy 1801 TCAAGGGGCGAGTGTCCCTGATCAATAGCTAATAATGGGACTTCTTCAACTACATTAAC 1860  
Dh 40445 TCAAGGGGCGAGTGTCCCTGATCAATAGCTAATAATGGGACTTCTTCAACTACATTAAC 40504  
Qy 1861 CAGAGGGCAAGACTTAATAATGCTGTAATGAACATTCGCTGTTCAATCTCAGCAGAGT 1920  
Dh 40505 CAGAGGGCAAGACTTAATAATGCTGTAATGAACATTCGCTGTTCAATCTCAGCAGAGT 40564  
Qy 1921 GTTGCAATTTGAGCTTCTGCAAGGGCCACCGAGACCTAATCTGCTCAGATGTTTAACTCA 1980  
Dh 40565 GTTGCAATTTGAGCTTCTGCAAGGGCCACCGAGACCTAATCTGCTCAGATGTTTAACTCA 40624  
Qy 1981 TCTAATTCAGTGAACACTTCA 2001  
Dh 40625 TCTAATTCAGTGAACACTTCA 40645

RESULT 8  
AX033911 LOCUS 40917 bp DNA linear PAT 21-SEP-2000  
DEFINITION Sequence 3 from Patent WO9851790.  
ACCESSION AX033911  
VERSION AX033911.1 GI:10280479  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1  
AUTHORS Cancelli, M.R., Choo, K.H., and Du, S.D.  
TITLE A novel nucleic acid molecule  
JOURNAL Patent: WO 9851790-A 3 19-NOV-1998;  
CANCELLI, MICHAEL ROBERT (AU) ; CHOO, KONG HONG ANDY (AU) ; SART

DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)  
Location/Qualifiers  
1. 40917  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 95.9%; Score 1918; DB 6; Length 40917;  
Best Local Similarity 100.0%; Pred.No.0;  
Matches 1918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGAATCACCAGCCTTGCCCTCCCAAGTGTGGATTATTAAGGCGTGAAGCCACCCGCTG 60  
Db 39000 TGAATCACCAGCCTTGCCCTCCCAAGTGTGGATTATTAAGGCGTGAAGCCACCCGCTG 39059  
QY 61 GTGGAATGCTTTATTTATTTGAAGACAACATGGCCCTTAAATCTGTCTTATTTGAC 120  
Db 39060 GTGGAATGCTTTATTTATTTGAAGACAACATGGCCCTTAAATCTGTCTTATTTGAC 39119  
QY 121 AGACTTTGATGAGTCAATCCCAATGCTGCCACTTACTGAACGGCCTTAATGACTTAC 180  
Db 39120 AGACTTTGATGAGTCAATCCCAATGCTGCCACTTACTGAACGGCCTTAATGACTTAC 39179  
QY 181 TCTCTCTGAGCTGTCTTCTGCAATATGTAAGTGAATATGATGGCTTCAAGGAGAA 240  
Db 39180 TCTCTCTGAGCTGTCTTCTGCAATATGTAAGTGAATATGATGGCTTCAAGGAGAA 39239  
QY 241 TAAACCTATGAAAAGTGTGAGGATAGTGTGATGTAATGAATTAAGATTTCAACAATAG 300  
Db 39240 TAAACCTATGAAAAGTGTGAGGATAGTGTGATGTAATGAATTAAGATTTCAACAATAG 39299  
QY 301 TAGCTGCTATGTAAGATTTAAGATTTATTTATTAACAATTTTAAATAAATTTAAAT 360  
Db 39300 TAGCTGCTATGTAAGATTTAAGATTTATTTATTAACAATTTTAAATAAATTTAAAT 39359  
QY 361 TAAATCACTTAATTTATTAAGAGCTTTGAATAGGCGCAAGGCTTACCTGCTGCTG 420  
Db 39360 TAAATCACTTAATTTATTAAGAGCTTTGAATAGGCGCAAGGCTTACCTGCTGCTG 39419  
QY 421 AATCCCAACATTTGGGAGGCGCAAGGTGGCGGATCACTGAGGTGAGGATTTAAGAC 480  
Db 39420 AATCCCAACATTTGGGAGGCGCAAGGTGGCGGATCACTGAGGTGAGGATTTAAGAC 39479  
QY 481 AGCTGCGCAACATGCTGTAACCCCTGTCTACTATAAAAACGCAAAATTTAGCCAGGTG 540  
Db 39480 AGCTGCGCAACATGCTGTAACCCCTGTCTACTATAAAAACGCAAAATTTAGCCAGGTG 39539  
QY 541 GTGGCATGACCTGTGATGCCAACTA CTGAGAGGTGAGGAGGAAATTTGCTTGAAC 600  
Db 39540 GTGGCATGACCTGTGATGCCAACTA CTGAGAGGTGAGGAGGAAATTTGCTTGAAC 39599  
QY 601 TAGAGGTGAGGATTCAGTAAACCCGAGATGCTACGCACTCCAGCCTGGCAACAGAGA 660  
Db 39600 TAGAGGTGAGGATTCAGTAAACCCGAGATGCTACGCACTCCAGCCTGGCAACAGAGA 39659  
QY 661 AGACTTCATTAAGACAACAAAGCTTTGAATTTGTAATGATGATGTAATCTTCAAT 720  
Db 39660 AGACTTCATTAAGACAACAAAGCTTTGAATTTGTAATGATGATGTAATCTTCAAT 39719  
QY 721 TTAAGAAATTCATCTTTGTCATTTATTTTACTTGAATGAGAGCTTCAAGAAATTTT 780  
Db 39720 TTAAGAAATTCATCTTTGTCATTTATTTTACTTGAATGAGAGCTTCAAGAAATTTT 39779  
QY 781 AATTAAGCCTCAACAGATTTTATGCTACGTATGATTAACAATTTATTTGCTAATA 840  
Db 39780 AATTAAGCCTCAACAGATTTTATGCTACGTATGATTAACAATTTATTTGCTAATA 39839  
QY 841 TAAATATCTTGTCTCTTTTAAAGAAATGCTCCTTAAGAAAGCTTTGTACCAAAAT 900  
Db 39840 TAAATATCTTGTCTCTTTTAAAGAAATGCTCCTTAAGAAAGCTTTGTACCAAAAT 39899  
QY 901 ACACTGACTTTACAAAATCAGATCTGATGTCACAAGTTGCAATGTTTTCAAAAAGAT 960

Db 39900 AACACTGACTTTACAAAATCAGATCTGATGTCACAAGTTGCAATGTTTTCAAAAAGAT 39959  
QY 961 TTTTCATTTGAAGAGGGGCCATTTGGGTATTTAGTTCTTAAGACTGAACGCTTTGT 1020  
Db 39960 TTTTCATTTGAAGAGGGGCCATTTGGGTATTTAGTTCTTAAGACTGAACGCTTTGT 40019  
QY 1021 TCTGTTTTCTGGCTCTGAGAGAGAGAGCATGAATTTCACTTGAACCTTTGATTT 1080  
Db 40020 TCTGTTTTCTGGCTCTGAGAGAGAGAGCATGAATTTCACTTGAACCTTTGATTT 40079  
QY 1081 TCTTTATCTTCAATTTCAATAGAGAGATGCTTCAATATGACAGTGTGTCAAT 1140  
Db 40080 TCTTTATCTTCAATTTCAATAGAGAGATGCTTCAATATGACAGTGTGTCAAT 40139  
QY 1141 CAAGAAAGAGAAACAGTTTCTGGTTTATTTTCAACCGGAAAGAAAGCAACCA 1200  
Db 40140 CAAGAAAGAGAAACAGTTTCTGGTTTATTTTCAACCGGAAAGAAAGCAACCA 40199  
QY 1201 TTTTGTGCGCTCTAATTTAGCCAGTCCATGACTTGAAGGCAAGGCTTTGAAGG 1260  
Db 40200 TTTTGTGCGCTCTAATTTAGCCAGTCCATGACTTGAAGGCAAGGCTTTGAAGG 40259  
QY 1261 GTGCTAACAGAGTCTTCAATATCTCCAGAGACTTCACTTCAATGCTAAGC 1320  
Db 40260 GTGCTAACAGAGTCTTCAATATCTCCAGAGACTTCACTTCAATGCTAAGC 40319  
QY 1321 TGAAGAAATGCTCAGAAAGATGAACAATCTCACAGACCCTTAACCTGAAGCCAG 1380  
Db 40320 TGAAGAAATGCTCAGAAAGATGAACAATCTCACAGACCCTTAACCTGAAGCCAG 40379  
QY 1381 TGTATTAAGCAACAATCAAGAGGTGAGAACTAACGTTCTTGAATCTCCCACTTCT 1440  
Db 40380 TGTATTAAGCAACAATCAAGAGGTGAGAACTAACGTTCTTGAATCTCCCACTTCT 40439  
QY 1441 TCTAGCTCAGAAAGGCAAGCTGATTTTATTTGTTGAATTTAAGAAATTTTAAATTA 1500  
Db 40440 TCTAGCTCAGAAAGGCAAGCTGATTTTATTTGTTGAATTTAAGAAATTTTAAATTA 40499  
QY 1501 TTCTAAGTCATGAACAAGCTTAATTTAAGATAGTTGCTGTAAGGTCTGAATAATAC 1560  
Db 40500 TTCTAAGTCATGAACAAGCTTAATTTAAGATAGTTGCTGTAAGGTCTGAATAATAC 40559  
QY 1561 TCGATTTTACCAACCCCTCTCTGAGAGAGCCATTAATGAAATCTGTGAACATGTTTAC 1620  
Db 40560 TCGATTTTACCAACCCCTCTCTGAGAGAGCCATTAATGAAATCTGTGAACATGTTTAC 40619  
QY 1621 TCTACCAAGAACTCTGTTTTTCTAATGAGAAACAGAGGCCCAAGATTTAACTATC 1680  
Db 40620 TCTACCAAGAACTCTGTTTTTCTAATGAGAAACAGAGGCCCAAGATTTAACTATC 40679  
QY 1681 TTAACCAATACAAATGACTAGTGTCTGTCTTTTATTAAGCACTAAATTTTATTC 1740  
Db 40680 TTAACCAATACAAATGACTAGTGTCTGTCTTTTATTAAGCACTAAATTTTATTC 40739  
QY 1741 AATTAATATCTGTCCATTTAAGAGATTTCCCTAATGTAAGTGTCTTAATCTTTCCT 1800  
Db 40740 AATTAATATCTGTCCATTTAAGAGATTTCCCTAATGTAAGTGTCTTAATCTTTCCT 40799  
QY 1801 TCAAGGGGCGAAGTCCGCTACACATAGCTAATGAGGACCTTCTCAACCTACATTAAC 1860  
Db 40800 TCAAGGGGCGAAGTCCGCTACACATAGCTAATGAGGACCTTCTCAACCTACATTAAC 40859  
QY 1861 CAGAGGCGAAGCTTAATAATGCTGTAATGACATTTCTGCTTCAACATCTCAGACCA 1918  
Db 40860 CAGAGGCGAAGCTTAATAATGCTGTAATGACATTTCTGCTTCAACATCTCAGACCA 40917

RESULT 9  
AC084084 71032 bp DNA linear HTG 25-MAR-2001  
LOCUS AC084084 Homo sapiens chromosome 8 clone RP11-314C19 map 8, LOW-PASS  
DEFINITION SEQUENCE SAMPLING.

ACCESSION AC084084  
VERSION AC084084.2 GI:13446278  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 71032)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, R.  
TITLE Homo sapiens chromosome 8, clone RP11-314C19  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 71032)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,  
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,  
Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,  
FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,  
Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,  
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRoque, K.,  
Lamzats, R., Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G.,  
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKenna, K.,  
McPheeters, R., Meldrim, J., Meneus, L., Minova, T., Mlenga, V.,  
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,  
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,  
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,  
Sougnuez, C., Spencer, B., Strange-Thomann, N., Stojanovic, N.,  
Straus, N., Subramanian, A., Talamas, J., Testa, J., Theodore, J.,  
Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,  
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
Zimmer, A. and Zody, M.  
DIRECT SUBMISSION  
Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 25, 2001 this sequence version replaced gi:10799449.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L11327  
Center clone name: 314\_C\_19  
-----  
\* NOTE: This record contains 85 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\*  
\* 1 720: contig of 720 bp in length  
\* 721 820: gap of 100 bp  
\* 821 1557: contig of 737 bp in length  
\* 1558 1657: gap of 100 bp  
\* 1658 2404: contig of 747 bp in length  
\* 2405 2504: gap of 100 bp  
\* 2505 3299: contig of 795 bp in length  
\* 3300 3399: gap of 100 bp  
\* 3400 4153: contig of 754 bp in length  
\* 4154 4253: gap of 100 bp  
\* 4254 4978: contig of 725 bp in length  
\* 4979 5078: gap of 100 bp  
\* 5079 5815: contig of 737 bp in length  
\*  
\* 5816 5915: gap of 100 bp  
\* 5916 6648: contig of 733 bp in length  
\* 6649 6748: gap of 100 bp  
\* 6749 7520: contig of 772 bp in length  
\* 7521 7620: gap of 100 bp  
\* 7621 8347: contig of 727 bp in length  
\* 8348 8448: gap of 100 bp  
\* 8449 9227: contig of 780 bp in length  
\* 9228 9328: gap of 100 bp  
\* 9329 10052: contig of 725 bp in length  
\* 10053 10152: gap of 100 bp  
\* 10153 10869: contig of 717 bp in length  
\* 10870 10969: gap of 100 bp  
\* 10970 11718: contig of 749 bp in length  
\* 11719 11818: gap of 100 bp  
\* 11819 12557: contig of 733 bp in length  
\* 12558 12657: gap of 100 bp  
\* 12658 13412: contig of 755 bp in length  
\* 13413 13512: gap of 100 bp  
\* 13513 14255: contig of 743 bp in length  
\* 14256 14355: gap of 100 bp  
\* 14356 15086: contig of 731 bp in length  
\* 15087 15186: gap of 100 bp  
\* 15187 15901: contig of 725 bp in length  
\* 15902 16001: gap of 100 bp  
\* 16002 16711: contig of 710 bp in length  
\* 16712 16811: gap of 100 bp  
\* 16812 17539: contig of 728 bp in length  
\* 17540 17639: gap of 100 bp  
\* 17640 18351: contig of 712 bp in length  
\* 18352 18451: gap of 100 bp  
\* 18452 19162: contig of 711 bp in length  
\* 19163 19262: gap of 100 bp  
\* 19263 20054: contig of 792 bp in length  
\* 20055 20154: gap of 100 bp  
\* 20155 20881: contig of 737 bp in length  
\* 20892 20991: gap of 100 bp  
\* 20992 21720: contig of 729 bp in length  
\* 21721 21820: gap of 100 bp  
\* 21821 22553: contig of 733 bp in length  
\* 22554 22653: gap of 100 bp  
\* 22654 23420: contig of 767 bp in length  
\* 23421 23520: gap of 100 bp  
\* 23521 24234: contig of 714 bp in length  
\* 24235 24334: gap of 100 bp  
\* 24335 25069: contig of 735 bp in length  
\* 25070 25169: gap of 100 bp  
\* 25170 25915: contig of 746 bp in length  
\* 25916 26015: gap of 100 bp  
\* 26016 26751: contig of 736 bp in length  
\* 26752 26851: gap of 100 bp  
\* 26852 27574: contig of 723 bp in length  
\* 27575 27674: gap of 100 bp  
\* 27675 28437: contig of 763 bp in length  
\* 28438 28537: gap of 100 bp  
\* 28538 29258: contig of 721 bp in length  
\* 29259 29358: gap of 100 bp  
\* 29359 30068: contig of 710 bp in length  
\* 30069 30168: gap of 100 bp  
\* 30169 30918: contig of 750 bp in length  
\* 30919 31018: gap of 100 bp  
\* 31019 31770: contig of 752 bp in length  
\* 31771 31870: gap of 100 bp  
\* 31871 32658: contig of 788 bp in length  
\* 32659 32758: gap of 100 bp  
\* 32759 33501: contig of 743 bp in length  
\* 33502 33602: gap of 100 bp  
\* 33603 34347: contig of 746 bp in length  
\* 34348 34447: gap of 100 bp  
\* 34448 35165: contig of 718 bp in length  
\* 35166 35265: gap of 100 bp  
\* 35266 35998: contig of 733 bp in length  
\* 35999 36098: gap of 100 bp





```

TITLE      Direct Submission
JOURNAL    Submitted (03-JUN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   7 (bases 1 to 122076)
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Submitted (10-JUN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8 (bases 1 to 122076)
AUTHORS    Wilson,R.K.
TITLE      Direct Submission
JOURNAL    Submitted (16-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT    On Nov 8, 2001 this sequence version replaced gi:15638775.
-----
            Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            Summary Statistics
            Center project name: H_NH0172J01
            -----

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPc1-11 human BAC library was made from the blood of one male
donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1996) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR:     pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-223M6, 200 bp overlap.
Actual start of this clone is at base position 1 of RP11-172J1;
actual end is at base position 33760 of RP11-223M6.
Location/Qualifiers
    source
    1..122076
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="2"
        /clone="RP11-172J1"
        /clone_1fb="RPc1-11"
        112630..112861
        /note="cpg_island (%GC=60.8, o/e=1.00, #cpgs=21)"

misc_feature
ORIGIN
Query Match      11.7%; Score 233.6; DB 8; Length 122076;
Best Local Similarity 61.9%; Pred. No. 1,8e-38;
Matches 422; Conservative 0; Mismatches 254; Indels 6; Gaps 3;

274 ATGTGAATATGAGATTTCACCAAGTAGTAGCTGCTATTTAGAGATTTAAGATTATTATT 333

```

Db	Accession	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
Db	22195	AAACAAAATACTATATAAAATTTATAAAATATATATATATAAAATATTTTAAATATTT	22133									
Qy	334	ACAACTATTTTAAATTAATTTTAAAACTATATACCTTAATTTATTTAAAGCTTTGAAT	393									
Db	22135	ATAATTAATTTATATAAAATTTATTAATAATTTATTAATAAGTATTTATAAATATATATTTTA	22078									
Qy	394	GGGCGAGGCGCACTACTCTCTGCTGTATATCCACAATTGGAGGCGCAAGGTGGCGG	453									
Db	22077	GGGCGAGAGTGTACTACACCTGTATATCCAGCATTTTGGAGGCGCAAGTGGCGAG	22018									
Qy	454	ATCACCTGAGTACAGAGATTAAAGACACACCTGGCCACATGGTGAACCCCTGTCTAC	513									
Db	22017	ATCACCTGAGTACAGAGATTCAAGTCCACCTGGCCACATGATGAACCCCATCTAC	21958									
Qy	514	TAAAAACGCAAAATTAAGCAGGTGTGTGGCATGACCTGTATGTCCTCACTACAGCA	573									
Db	21957	TAAAAATATATAAAATTAATCACTGGGTGTGGGCAACACCTGTATTTCCAGCTACAGCA	21898									
Qy	574	GGTTGAGGAGGAGGAATTTGCTTAACCTTGAAGGTGAGGTGAGGTGACATGCCGAG	630									
Db	21897	GGGTGAGGAGGAGGAATTCGCTTAAACCCAGAGGTGTGAGGTGTGCGCAGCTGATCAT	21838									
Qy	631	GTCACTGCACTCCAGCT-GGCAACAGAGCAAGACTCCATTAAGAACAACAAAGCTTGA	689									
Db	21837	GCCACTGCACTCCAGCTGGTGGCAACAGAGAGACTCCATTCACAAAAAATAATGA	21778									
Qy	690	AATTGTGTAATGAGTTGTAACCTATCTTCAATTAAGAAATTCATCTTTGTATATATT	749									
Db	21777	CTTGTTATATCTTTTCTTAATCCCTTAATTTGAATTTTCAGTGTATTAATTTAGGTGTTT	21718									
Qy	750	TTACTTGAACATGAGAGCTTCAGCAATTTTAAATTAAGCCCTCACAGATTTATGTCAC	809									
Db	21717	TCTTAATATATGCTGAATTAATAAAACATCTTATTTATCTTTATATAGTATTTTAAAC	21658									
Qy	810	GGCTATGTAATTAACAATATATTTTGCTAAATAATATTTCTTGCTTTTAAAGAAAT	869									
Db	21657	CTTATATATTTGTTGTTACTACTGTTAATTAATAAAACATTTTACTTTATATACAGATAT	21598									
Qy	870	GTCCTCCTAGAAACGGTTGTACCAACAATACACTCTTACACAAATCAGATCTGA	929									
Db	21597	TTTCTCTTACTCTCCCTACTCCCTGGTTATTAATCTTGATGATCTTTTCTCAATTTT	21538									
Qy	930	TTGGCAACAGTTGCAGATGTTT	951									
Db	21537	TTTCTCTCTGGTTGAAGATTT	21516									
RESULT 11	AC144747/c	140842 bp	DNA	linear	HTG 07-JUN-2003							
LOCUS	AC144747	140842 bp	DNA	linear	HTG 07-JUN-2003							
DEFINITION	Pan troglodytes clone CH251-160J7, WORKING DRAFT SEQUENCE, 5											
ORDERED	ordered pieces.											
ACCESSION	AC144747											
VERSION	AC144747.2	GI:31455634										
KEYWORDS	HTG, HTGS, PHASE2, HTGS, DRAFT.											
SOURCE	Pan troglodytes (chimpanzee)											
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Pan.											
REFERENCE	1 (bases 1 to 140842)											
AUTHORS	Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Cariga, K., Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X., Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-D., Hu, P., Hurle, B., Idol, J.R., Karlins, E., Kwong, P., Larcic, P., Lee-Jin, S.-O., Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Mastello, C., Maskeri, B., McDowell, J., Nagurigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Reddi-Dugue, N., Schandler, K., Schneider, M.G., Shan, K., Sison, C., Stantirpop, S., Thomas, J.W., Thomas, P.D., Tsipouris, V., Vogt, J.L., Weherby, K.D., Wiggins, L., Young, A. and Green, E.D.											
TITLE	NISC Comparative Sequencing Initiative											

REFERENCE 2 (bases 1 to 140842)  
AUTHORS Green, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (15-MAY-2003) NIH Intramural Sequencing Center, 8717  
Groveomont Circle, Gaithersburg, MD 20877, USA  
REFERENCE 3 (bases 1 to 140842)  
AUTHORS Green, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (07-JUN-2003) NIH Intramural Sequencing Center, 8717  
Groveomont Circle, Gaithersburg, MD 20877, USA  
On Jun 7, 2003 this sequence version replaced gi:30725904.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc\_zoo@nhgri.nih.gov  
----- Project Information  
Center project name: esb  
Center clone name: 160D07

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 140151 bases at least Q40  
Consensus quality: 140382 bases at least Q30  
Consensus quality: 140426 bases at least Q20  
Insert size: 13200; agarose-fp  
Insert size: 140442; sum-of-contigs  
Quality coverage: 11.84x in Q20 bases; agarose-fp  
Quality coverage: 11.13x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 60845: contig of 60845 bp in length  
\* 60846 60945: gap of unknown length  
\* 60946 99599: contig of 38654 bp in length  
\* 99600 99699: gap of unknown length  
\* 99700 102357: contig of 2558 bp in length  
\* 102358 102357: gap of unknown length  
\* 102358 128790: contig of 26433 bp in length  
\* 128791 128890: gap of unknown length  
\* 128891 140842: contig of 11952 bp in length.

FEATURES  
Source  
1. .140842  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="CH251-160D7"  
/clone\_lib="CH251"

misc\_feature  
1. .101605  
/note="clone overlaps with GenBank Accession Number AC144872 clone CH251-495114 (center project name esa)"

misc\_feature  
1. .60845

/note="assembly\_fragment  
clone\_end:T7  
vector\_side:left"  
60846..60945  
/estimated\_length=unknown  
60946..99599  
/note="assembly\_fragment"  
84997..140842  
/note="clone overlaps with GenBank Accession Number AC096878 clone RP43-171C12 (center project name clis)"  
99600..99699  
/estimated\_length=unknown  
99700..102357  
/note="assembly\_fragment"  
102358..102357  
/estimated\_length=unknown  
102358..128790  
/note="assembly\_fragment"  
128791..128890  
/estimated\_length=unknown  
128891..140842  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:right"

misc\_feature  
128891..140842

gap  
/estimated\_length=unknown  
99700..102357  
/note="assembly\_fragment"

misc\_feature  
102358..128790  
/note="assembly\_fragment"

gap  
/estimated\_length=unknown  
128791..128890

misc\_feature  
128891..140842

ORIGIN

Query Match 11.2%; Score 224.8; DB 14; Length 140842;  
Best Local Similarity 73.3%; Pred. No. 1.1e-36;  
Matches 302; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 377 TTAAGAGCTTTGAATGGCCAGGCGCAGTGTCTCCGCTGATATCCACATCTTNG 436  
DB 4056 TTAAGAGGAGGAGAAATGGCCAGGCGTCACTGCTCACTGCTGATATCCACATCTTNG 3997

QY 437 GAGGCCAAGGTGGCGGATCACTAGGTGAGAGATTGAAGACGAGCTGGCAACATGG 496  
DB 3996 GAGGCCAAGGTGGCGGATCACTAGGTGAGAGATTGAAGACGAGCTGGCAACATGG 3937

QY 497 TGAACCCCTGTCTCTAATAAAGCAAAATTAAGCAGGTGTGTGATCACTCTGA 556  
DB 3936 TGAACCCCTGTCTCTAATAAATTAAGCAGGTGTGTGATCACTCTGA 3877

QY 557 GTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAACCTAGAGGTGAGTTG 616  
DB 3876 ATCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAACCTAGAGGTGAGTTG 3817

QY 617 CAGTACCCGAGAT---GTCACTGCACTCCAGCTGGCAAGAGACTCCATTAAG 673  
DB 3816 CAGTACCCGAGATCGCGTCACTGCACTCCAGCTGGCAAGAGACTCCATTAAG 3757

QY 674 ACAACAAAAGCTTGAATGTGTAATGATGTTGACTTATCTTCAATTAAGAAATTCAT 733  
DB 3756 AAAAAGAAAAGAAAGAAATTTGGCATGCTGTAACATTAAGAAATTCATTAAG 3697

QY 734 CTTTGTCTATTATTTTACTTGAATGAGAGCTTCAGACTTTTATTAATTA 785  
DB 3696 CTTTGTCTATTATTTTACTTGAATGAGAGCTTCAGACTTTTATTAATTA 3645

RESULT 12  
AC144872/c 173623 bp DNA linear HTG 07-JUN-2003  
LOCUS Pan troglodytes clone CH251-495114, WORKING DRAFT SEQUENCE, 2  
DEFINITION  
ordered pieces.  
ACCESSION  
AC144872  
VERSION  
AC144872.2 GI:31455635  
KEYWORDS  
HTG; HTGS PHASE2; HTGS DRAFT.  
SOURCE  
Pan troglodytes (chimpanzee)  
ORGANISM  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Pan.  
REFERENCE  
1 (bases 1 to 173623)

AUTHORS Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Cariega, K., Chu, G., Coleman, B., Coleman, H., Engle, J., Grant, S., Guan, X., Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P., Hurle, B., Idol, J.R., Karlins, E., Kwong, P., Latic, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masiello, C., Mazeri, B., McDowell, J., Pagutigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Reddix-Dugue, N., Schandler, K., Schueler, M.G., Shah, K., Sison, C., Stanciripop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V., Vogt, J.L., Weherby, K.D., Wiggins, L., Young, A. and Green, E.D.

TITLE NISC Comparative Sequencing Initiative

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 173623)

Green, E.D.

Direct Submission

Submitted (23-MAY-2003) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Galtersburg, MD 20877, USA

3 (bases 1 to 173623)

Green, E.D.

Direct Submission

Submitted (07-JUN-2003) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Galtersburg, MD 20877, USA

On Jun 7, 2003 this sequence version replaced gi:31044285.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: [nisc\\_zoo@nhgri.nih.gov](mailto:nisc_zoo@nhgri.nih.gov)

----- Project Information

Center project name: esa

Center clone name: 495114

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 173453 bases at least Q40

Consensus quality: 173497 bases at least Q30

Consensus quality: 173518 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 173523; sum-of-contigs

Quality coverage: 13.17x in Q20 bases; agarose-fp

Quality coverage: 11.92x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 132061: contig of 132061 bp in length

\* 132062 132161: gap of unknown length

\* 132162 173623: contig of 41462 bp in length.

Location/Qualifiers

1. 173623

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

FEATURES

source

misc\_feature /clone="CH251-495114" /clone\_id="CH251" 1..132061 /note="assembly\_fragment clone end:17 vector\_side:left"

misc\_feature 1..65891 /note="clone overlaps with GenBank Accession Number AC096851 clone RP43-1961 (center project name clt)" 71148..173623 /note="clone overlaps with GenBank Accession Number AC144747 clone CH251-160D7 (center project name esb)"

gap 132062..132161 /estimated\_length=unknown

misc\_feature 132162..173623 /note="assembly\_fragment clone end:SP6 vector\_side:right"

misc\_feature 156029..173623 /note="clone overlaps with GenBank Accession Number AC096878 clone RP43-171C12 (center project name clb)"

ORIGIN

Query Match 11.2% Score 224.8; DB 14; Length 173623;

Best Local Similarity 73.3% Pred. No. 1.1e-36;

Matches 302; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 377 TTAAGAGCTTTGAATGGGCGGCGAGTACCTGCTGCTGTAATCCCACTTTGG 436

DB 75209 TTTAAAGAGGAGAAATTTGGCCAGGCTCAGTGGCTGATATCCACACTTTGG 75150

QY 437 GAGGCCAAGTGGGCGGATCACCTGAGTCAAGATTAAAGCCAGCTGGCCACATGG 496

DB 75149 GAGGCCAAGTGGGCGGATCACCTGAGTCAAGATTAAAGCCAGCTGGCCACATGG 75090

QY 497 TGAACCTGCTCTACTTAAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAG 556

DB 75089 TGAACCTGCTCTACTTAAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAG 75030

QY 557 GTCCCAACACTACAGGAGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 616

DB 75029 ATCCCACTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 74970

QY 617 CAGTAACCCGAGAT--GTCACTGCACTCCAGCTGGCAACAGAGCAATCTTCATTAAG 673

DB 74969 CAGTAACCCGAGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 74910

QY 674 ACAACAAAGCTTTGAATTTGTGAATTTGTGAATTTGTGAATTTGTGAATTTGTGAATTT 733

DB 74909 AAAAAGAAAAGAAAGAAATTTGGCATGCTGTAACATTAAGAACTGGAGGCAATTATG 74850

QY 734 CTTTGTTCATTTATTTTAACTTGACATGAGAGCTTCAGCAATTTTAAATTA 785

DB 74849 CTGAGTGAATTAAGCCAGTCAATTTTAACTTGACATGAGAGCTTCAGCAATTTTAAATTA 74798

RESULT 13

AC121247 170787 bp DNA linear PRI 19-NOV-2002

LOCUS AC121247

DEFINITION Homo sapiens chromosome 3 clone RP11-3B7, complete sequence.

AC121247 AC012124

VERSION AC121247.2 GI:25101130

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 170787)

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphitachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.

TITLE Direct Submission

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 170787)  
 AUTHORS Kaul, R. K., Olson, M. V., Raymond, C. and Haugen, E. D.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAY-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 REFERENCE 3 (bases 1 to 170787)  
 AUTHORS Kaul, R. K., Olson, M. V., Zhou, Y., James, R. A., Rouse, G., Wu, Z.,  
 Saephammachak, C., Buckley, D., Kibukawa, M., Raymond, C. and  
 Haugen, E. D.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-NOV-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 COMMENT On Nov 19, 2002 this sequence version replaced gi:20806311.

Center: University of Washington Genome Center  
 Center Code: UMG  
 Web site: <http://www.genome.washington.edu>  
 Contact: [uwgchtg@u.washington.edu](mailto:uwgchtg@u.washington.edu)  
 Drafting Center: BCM

Project Information  
 Center project name: Chr-3  
 Center clone name: RP11-3B7 (bc0105)

Summary Statistics  
 Sequencing vector: plasmid; 47% of reads  
 Sequencing vector: unknown; 53% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 170463 bases at least Q40  
 Consensus quality: 170727 bases at least Q30  
 Consensus quality: 170782 bases at least Q20  
 Insert size: 170787; sum-of-contigs  
 Quality coverage: 9.4x in Q20 bases; sum-of-contigs

Overlapping Sequences:  
 5': Mapping in progress  
 3': RP11-804H15 AC104311

Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII EcoRI BglII

SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

13802	13540	8696	8757	3877	3864
6382	6542	6	<800	2067	2046
512	<800	9988	9979	14311	14181
449	<800	8347	8314	2993	3053
5288	5288	1772	1747	3829	3864
660	<800	12907	12726	2535	2519
835	854	929	930	3072	3053
7851	7883	8700	8757	10768	10611
3365	3374	693	<800	1072	1062
39	<800	2756	2797	39	<800
4109	4061	14794	14646	5511	5565
189	<800	12996	12726	4883	4732
1697	1649	1165	1149	15360	15166
15845	15867	1309	1284	110	<800
941	948	3194	3202	3439	3511
5249	5288	7412	7507	1558	1529
8491	8664	15188	14646	769	778
3057	3069	13287	12726	3706	3864
4457	4471	3102	3087	4759	4732
1342	1306	4166	4243	212	<800
9716	9838	1452	1420	4701	4732
24363	24680	794	806	5330	5368
3172	3183	3383	3398	586	<800
10795	10782	2234	2257	1740	1716
4741	4684	3035	3087	37	<800
21198	21101	3377	3398	1272	1241
4653	4684	10041	9979	3479	3511
16285	15867	578	<800	678	<800
		10129	9979	5235	5176
		13053	12726	7875	8032
				22885	23142
				12447	12320
				5515	5565
				4945	5004
				7736	7788
				10152	10081

FEATURES

Location/Qualifiers

## SOURCE

1. 170787  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /clone="RP11-387"  
 /clone\_lib="RP11 human BAC library 11"

## ORIGIN

Query Match 11.2%; Score 224.4; DB 8; Length 170787;  
 Best Local Similarity 79.1%; Pred. No. 1.3e-36;  
 Matches 280; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 334 ACACTATTATTAATAATTTTAAACACTATATCACTTAATATTAAGAGCTTTGAAT 393  
 DB 138920 ACAAGAGTAAATCTCATCTCGAAAAAATATATATTAATTAATTAATTAATTA 138979  
 QY 394 GGGCCAGGCGAGTACCTCTGCTGATCCCAACACTTTGGAGGCCAAGTGGCGG 453  
 DB 138980 TGGCCAGGCGAGTACCTCTGCTGATCCCAACACTTTGGAGGCCAAGTGGCGG 139039  
 QY 454 ATCACTGAGTACAGAGTTTAAAGACAGCTGCGCCACATGATGTAACCTGTCTTAC 513  
 DB 139040 ATCACTGAGTACAGAGTTTAAAGACAGCTGCGCCACATGATGTAACCTGTCTTAC 139099  
 QY 514 TAAAAACGAAATATGCGAGCTGTGGCATGCACTTGTCCCACTACTACAGA 573  
 DB 139100 TAAAAATACAAATATGCGAGCATGTGGCGAGGCGCTGTATGCTCCAGTACTAGGA 139159  
 QY 574 GGTGAGGAGAGAGATTCCTTGAACCTTGGAGGTGAGGTGACATACCGAGA--T 630  
 DB 139160 GGCTGAGGAGAGAGATTCCTTGAACCTTGGAGGTGAGGTGACATACCGAGA--T 139219  
 QY 631 GTCACTGCACTCGAGCTGCGCAACAGAGCAAGATCCATTAAGACAAACAAAGC 684  
 DB 139220 GCCACTGCACTCGAGCTGCGCAACAGAGCAAGATCCATTAAGACAAACAAAGC 139273

## RESULT 14

AC135506/c 174363 bp DNA linear PRI 23-APR-2003  
 LOCUS Homo sapiens chromosome 3 clone RP11-694115, complete sequence.  
 DEFINITION AC135506  
 AC135506  
 VERSION AC135506.3 GI:30061452  
 KEYWORDS HTG.  
 SOURCE  
 ORGANISM Homo sapiens (human)

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 174363)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Senphimachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.  
 Direct Submission

REFERENCE  
 AUTHORS Unpublished  
 2 (bases 1 to 174363)  
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
 Direct Submission  
 Submitted (17-OCT-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 174363)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Senphimachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.  
 Direct Submission

REFERENCE  
 AUTHORS Submitted (01-MAR-2003) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 4 (bases 1 to 174363)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Senphimachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.  
 Direct Submission

## JOURNAL

## COMMENT

Submitted (23-APR-2003) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 On Apr 23, 2003 this sequence version replaced gi:28626666.  
 -----  
 Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UWGC  
 Web site: <http://www.genome.washington.edu>  
 Contact: uwgchgs@u.washington.edu  
 -----  
 Project Information  
 Center project name: chr-3  
 Center clone name: RP11-694115 (bc0806)  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 174180 bases at least Q40  
 Consensus quality: 174339 bases at least Q30  
 Consensus quality: 174362 bases at least Q20  
 Insert size: 174363; sum-of-contigs  
 Quality coverage: 11.7x in Q20 bases; sum-of-contigs

## Overlapping Sequences:

5': RP13-131K19 AC137630, 2005-bp overlap  
 3': RP11-387 (UWGC:bc0105) AC121247, 33945-bp overlap

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HandIII	ECORI	BglII
SegDerMap FngPrnt	SegDerMap FngPrnt	SegDerMap FngPrnt
-----	-----	-----
3328	3315	8696
6382	6511	6
512	<800	929
449	<800	12907
11958	11682	1772
1697	1644	8347
189	<800	9988
		9926
		3829
		3822

4109	4017	1091	1071	2993	3010
39	<800	14963	15082	20894	21024
3355	3315	1661	1625	9237	8982
7851	7831	11265	11328	8585	8982
835	881	10108	9926	793	<800
660	<800	15190	15082	5462	5408
8679	8754	189	<800	648	<800
14244	14033	3890	3920	316	<800
5759	5860	5708	5766	9849	9784
4655	4677	3039	3054	1211	1211
6520	6511	15804	16006	1878	1881
8654	8754	11492	11328	433	<800
8623	8754	15945	16006	3907	3999
8927	8754	51	<800	20	<800
3967	4017	2131	2148	4872	4903
2675	2684	4567	4606	6115	6167
16757	16905	779	778	4033	3999
7172	7286	11477	11328	2184	2182
5547	5584	302	<800	4448	4483
801	828	9721	9926	868	955
1353	1336	1044	1071	920	955
447	<800	2598	2650	839	8982
7394	7510	839	8982	645	<800
3446	3448	645	4903	8855	8982
6132	6192	4731	4903	104	<800
890	881	8855	8982	978	955
8111	8115	104	<800	12456	12174
50	<800	978	955	2400	2373
1028	1028	12456	12174	4969	4903
1797	1774	2400	2373	5012	4903
4203	4017	4969	4903	175	<800
3847	3805	5012	4903	5839	5788
		175	<800	1080	1070

FEATURES  
source Location/Qualifiers  
1..174363

```
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-694I15"
/clone_lib="RP11 human BAC library 11"
140036..140049
/notes="single subclone region"
174350..174363
/notes="single subclone region"

ORIGIN
Query Match 11.2%; Score 224.4; DB 8; Length 174363;
Best Local Similarity 79.1%; Pred. No. 1.3e-36;
Matches 280; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 334 ACACCTATTATTAATAATTTTAAAACTAATACCTTAATTAATTAAGAGCTTTGAAT 333
Db 172282 ACAGAGTGAACCTCATCTCAAGAAAAATATATTAATTAATTAATTAATTA 172223

QY 394 GGGCCAGGCGCAGTACCTCTGCTGTATCCCAACCTTGGAGGCCAAGTGGCGG 453
Db 172222 TGGCCAGGCAAGTGGCTCAAGCTTAATTCAGCATTTTGGAGGCTGAAGTGGTG 172163

QY 454 ATCACTGAGTCAAGAGTTTAAAGACCAAGCTTGGCAATGGTGAACCCCTGCTTAC 513
Db 172162 ATCACTGAGTCAAGAGTTTGAACCAAGCTTGGCAATGGTGAACCCCTGCTTAC 172103

QY 514 TAAAAAGCAAAATTTAGCCAGGTGGTGGCAATGCACTGTAGTCCCACTACTAGGA 573
Db 172102 TAAAAATACAAAATTTAGCCAGCATGGTGGCAATGGTGAACCCCTGCTTAC 172043

QY 574 GGTGAGGAGAGAAATTCCTTGAACCTTGAAGAGTGGAGTGGCAATCCGAGA---T 630
Db 172042 GGCTGAGGAGAGAAATTCCTTGAACCTTGAAGAGTGGAGTGGCAATCCGAGA 171983

QY 631 GTCACTGCACTCAAGCTGGCAAGAGCAAGCAAGTCCATTAAGACAACAAAGC 684
Db 171982 GCCACTGCACTCAAGCTGGCAAGAGCAAGCAAGTCCATTAAGACAACAAAGC 171929

RESULT 15
AC104311/c 185146 bp DNA linear HTG 21-JAN-2002
LOCUS Homo sapiens chromosome UNK clone RP11-804H15, *** SEQUENCING IN
DEFINITION PROGRESS ***, 45 unordered pieces.
ACCESSION AC104311
VERSION AC104311.2 GI:18252763
KEYWORDS HTG; HTGS PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE 1 (bases 1 to 185146)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185146)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jan 21, 2002 this sequence version replaced gi:17402818.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H_NH0804H15
----- Summary Statistics -----
```





Query Match  
Best Local Similarity 11.2%; Score 224.4; DB 14; Length 185146;  
Matches 280; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

gap /note="assembly\_name:Contig115"  
13980..14079 /estimated\_length=unknown  
misc\_feature 14080..15503 /note="assembly\_name:Contig116"  
gap 15504..15603 /estimated\_length=unknown  
misc\_feature 15604..17771 /note="assembly\_name:Contig117"  
gap 17772..17871 /estimated\_length=unknown  
misc\_feature 17872..19412 /note="assembly\_name:Contig118"  
gap 19413..19512 /estimated\_length=unknown  
misc\_feature 19513..22030 /note="assembly\_name:Contig119"  
gap 22031..22130 /estimated\_length=unknown  
misc\_feature 22131..23419 /note="assembly\_name:Contig120"  
gap 23420..23519 /estimated\_length=unknown  
misc\_feature 23520..25351

QY 334 ACAACTATTAAATTTAAATTTAAATTAATACCTTAATTAATTAAGAGCTTTGAAT 393  
DB 120915 ACAAGTGAATCTCATCTCGAATAAATAATATTAATTAATTAATTAATTA 120856

QY 394 GGGCCAGCGCCAGTACTCTCTGCTGTATCCCAACATTTGGAGGCCAAGGTGGAG 453  
DB 120855 TGGCCAGGACAGTGTCTCAAGCTTAATTTCCAGATTTGGAGGCTGAGATGGGTG 120796

QY 454 ATCACTGAGTCAAGAGTTTAAAGACAGCTGGCCCAACATGTTGAACCTCTCTCTAC 513  
DB 120795 ATCACTGAGTCAAGAGTTTGAACAGCTGGCCCAACATGTTGAACCTCTCTCTAC 120736

QY 514 TAAAAACGAAAAATTAGCCAGGTGTGTGCATGACCTGTATCCCACTACTGAG 573  
DB 120735 TAAAAATACAAAAATTAGCCAGGCATGTGTGCAGGGCCGTATGCTCCAGCTACTGAG 120676

QY 574 GGTGAGGAGAGAAATTCCTTGAACCTTAGAGAGGTGAGGTTGCAGTAAACCGAGA--T 630  
DB 120675 GGCTGAGGAGAGAAATTCCTTGAACCTTAGAGAGGTGAGGTTGCAGTAAACCGAGA 120616

QY 631 GTCACTGCACTCAGCTGGCAACAGAGCAAGCTCCATTAAGACAAACAAAC 684  
DB 120615 GGCACCTGCACTCAGCTGGCAACAGAGCAAGCTCCATTAAGACAAACAAAC 120562

RESULT 16  
AC073546/c 141521 bp DNA linear PRI 21-JUL-2001  
LOCUS AC073546 Homo sapiens chromosome 5 clone RP11-406P13, complete sequence.  
DEFINITION AC073546  
VERSION AC073546.5 GI:14993704  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 141521)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 141521)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission

JOURNAL Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint  
REFERENCE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
AUTHORS 3 (bases 1 to 141521)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
REFERENCE Submitted (26-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
AUTHORS Drive, Walnut Creek, CA 94598, USA  
TITLE 4 (bases 1 to 141521)  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
COMMENT Direct Submission  
Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Jul 21, 2001 this sequence version replaced gi:14550307.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.  
Location/Qualifiers  
1..141521  
source /organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-406P13"

ORIGIN  
Query Match  
Best Local Similarity 11.2%; Score 224.2; DB 8; Length 141521;  
Matches 446; Conservative 0; Mismatches 278; Indels 15; Gaps 4;

QY 1 TGAATCAGCAGCTTGCGCTCCCAAGTGTCTGAGATTAACAGCGTGAAGCCAGCGCTG 60  
DB 91558 TGAATCAGCTTGCGCTCCCAAGTGTCTGAGATTAACAGCGTGAAGCCAGCGCTG 91499

QY 61 GTCAAGTGTCTTATTAATTTGAAGACAACATGGGCTTAAATCTGTCTTATTTGAC 120  
DB 91498 GCCATAGTCTTATTAATTTGAAGACAACATGGGCTTAAATCTGTCTTATTTGAC 91448

QY 121 AGACTTGTATGAGTCAAAATCCCAATGCTGCCACTTACTGAACGGCTTAATGACTTAG 180  
DB 91447 AAATTTATTCATATATCAACATCAACATTAAGACAGGCAAGAACTATATATCTGTA 91388

QY 181 TCTCTCAGCTGTCTTCTGCATATGTAAAGTGAATATGATGCTTTCAAGAGAA 240  
DB 91387 TTAATTTGTGTGTTTAAACCTTTTGTCTGATTTTATCAATTTTAAATGAAA 91328

QY 241 TAACTATGAAGAGTGTGAGATAGTGTGATATGAATTAAGATTCAACAAGTAG 300  
DB 91327 CTGGAAATTTACAGTACTGATTTGCTATTTGATGACATGCAAAATTTCTGTGACTATGT 91268

QY 301 TAGCTGCTATTGAAGATTTAAGATTTATTTATTAACAAT -ATTATTAATTTTAA 358  
DB 91267 GGAAGATTTCTTAGACTTAAGATTAAGATTTAAGATTTTCAACAATTAAGTTTCA 91208

QY 359 ACTAATACCTTAATTTAAGAGCTTTGAATAGGCGCAGGCGAGTACTCTGCT 418  
DB 91207 TGAATTTCAATTTAATTAATTAAGATTTAAGATTTGAGGCGAGGCTGAGGCT 91148

QY 419 GTATCCCAACCTTTGGAGGCAAGGAGGCGGATCACTGAGGTGAGAGTTTAA 478  
DB 91147 GTATCCCAACCTTTGGAGGCGGAGGAGGATCACTGAGGTGAGAGTTTAA 91088

QY 479 CCAGCTGGCCCAACATGTGAACCTGCTCTACTTAAACGCAAAATTTAGCCAGGTG 538  
DB 91087 CCAGCTGGCCCAACATGTGAACCTGCTCTACTTAAACGCAAAATTTAGCTGAGT 91028

QY 539 TGGTGCATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATTTGTTGA 598  
DB 91027 TGGTGCATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATTTGTTGA 90968

QY 599 CCTAGAGGTGAGAGTTTGAAGTAAACCGAG--ATGTCACTGCACTCCAGCT -GGCAAC 654

Db 90967 CCCAGAGGACAGAGTTGACGAGCAAGATCATCCACTGACCTTGAGCTGGGCAAC 90908  
QY 655 AGAGCAAGACTCCATTAAGACACAAACCTTGAAATGTGTAAATGCTGTACCTT 714  
Db 90907 AGAGCAAGACCTGCTGTGAAAAATTAAATTTAAAAAATTGGGTATTATTT 90848  
QY 715 CTTCAATTAAAGAAATTCAT 733  
Db 90847 AGTAACATTTTAAAGTAAT 90829

RESULT 17  
AC026465/c  
LOCUS AC026465 142838 bp DNA linear HTG 26-SEP-2000  
DEFINITION Homo sapiens chromosome 5 clone RP11-354J24, WORKING DRAFT  
SEQUENCE AC026465, 22 unordered pieces.  
ACCESSION AC026465  
VERSION AC026465.2 GI:10305170  
KEYWORDS HTG: HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
1 (bases 1 to 142838)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 142838)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Sep 26, 2000 this sequence version replaced gi:7280277.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 542541  
Center clone name: RPCI-11\_354J24  
-----  
Summary Statistics  
Consensus quality: 123712 bases at least Q40  
Consensus quality: 135403 bases at least Q30  
Consensus quality: 137509 bases at least Q20  
Estimated insert size: 194240; agarose-fp estimation  
Estimated insert size: 140738; sum-of-contigs estimation  
Quality coverage: 3.28 in Q20 bases; agarose-fp estimation  
Quality coverage: 4.53 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1225: contig of 1225 bp in length  
1226 1325: gap of unknown length  
1326 2352: contig of 1027 bp in length  
2352 2452: gap of unknown length  
2452 3799: contig of 1347 bp in length  
3799 3899: gap of unknown length  
3899 5202: contig of 1303 bp in length  
5202 5303: gap of unknown length  
5303 6648: contig of 1346 bp in length  
6648 6749: gap of unknown length  
6749 9853: contig of 3105 bp in length  
9853 9954: gap of unknown length  
9954 11640: contig of 1687 bp in length

\* 11641 11740: gap of unknown length  
\* 11741 13774: contig of 2034 bp in length  
\* 13775 13874: gap of unknown length  
\* 13875 15777: contig of 1903 bp in length  
\* 15778 15877: gap of unknown length  
\* 15878 17736: contig of 1859 bp in length  
\* 17737 17836: gap of unknown length  
\* 17837 20854: contig of 3018 bp in length  
\* 20855 20954: gap of unknown length  
\* 20955 24018: contig of 3064 bp in length  
\* 24019 24119: gap of unknown length  
\* 24119 29675: contig of 5557 bp in length  
\* 29676 29775: gap of unknown length  
\* 29776 36388: contig of 6613 bp in length  
\* 36389 43469: gap of unknown length  
\* 43470 43569: gap of unknown length  
\* 43570 51109: contig of 7540 bp in length  
\* 51110 51209: gap of unknown length  
\* 51210 61408: contig of 10199 bp in length  
\* 61409 71666: gap of unknown length  
\* 71667 85149: gap of unknown length  
\* 85150 85249: gap of unknown length  
\* 85250 100523: contig of 15274 bp in length  
\* 100524 119207: gap of unknown length  
\* 119208 119308: contig of 18584 bp in length  
\* 119308 142838: contig of 23531 bp in length.  
Location/Qualifiers  
1. 142838  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone\_1fb="RP11-354J24"  
/clone\_1fb="RPCI human BAC library 11"  
1226 .1325  
/estimated\_length=unknown  
2353 .2452  
/estimated\_length=unknown  
3800 .3899  
/estimated\_length=unknown  
5203 .5302  
/estimated\_length=unknown  
6649 .6748  
/estimated\_length=unknown  
9854 .9953  
/estimated\_length=unknown  
11641 .11740  
/estimated\_length=unknown  
13775 .13874  
/estimated\_length=unknown  
15778 .15877  
/estimated\_length=unknown  
17737 .17836  
/estimated\_length=unknown  
20855 .20954  
/estimated\_length=unknown  
24019 .24118  
/estimated\_length=unknown  
29676 .29775  
/estimated\_length=unknown  
36389 .36488  
/estimated\_length=unknown  
43470 .43569  
/estimated\_length=unknown  
51110 .51209  
/estimated\_length=unknown  
61409 .61508  
/estimated\_length=unknown  
71667 .71766

gap	/estimated_length=unknown 85150.85249
gap	/estimated_length=unknown 100524.100623
gap	/estimated_length=unknown 119208.119307
gap	/estimated_length=unknown

Query Match	11.2%;	Score 224.2;	DB 14;	Length 142838;
Best Local Similarity	60.4%;	Pred. No. 1.5e-36;		
Matches 446;	Conservative	0;	Mismatches 278;	Indels 15;
				Gaps 4;

OY	1	TGATCCACAGGCTTGCGCTCCCAAGTCTGGATTAACGGGTAGCACCAGCGCTG	60
Db	108011	TGATCTGCGCTGCTACGCTCTGMAAGTCTGGATTAACGGGTAGCACCAGC	1079532
OY	61	GTGGAATGTCCTTATTTATTTGAAGACAACATGGCGCTTAATCTGTCCTTATTTGAC	120
Db	107951	GCCATTAAGCTTTTAAAAACAATCTG-----TTAGCCACACTTAATGT	1079010
OY	121	AGACTTGAATGGAATCAATCCCAATGCTGCCCTTACTGAAAGCGCTTAATGACTT	180
Db	107900	AAAAATATATCATATATACCAACCATTAAGACAGGGCAGGAAATCTATATATATCTGTAA	1078411
OY	181	TCTCTCGACTGTCCTTTGCAATATGTAAGTGGAAATATGATGAGCTTTCAAGAGGAA	240
Db	107840	TTATTTTGTGGTTTAACTTTTAACTTTTGTCTTACGATTTTATTAATTTTAAATGAAA	1077811
OY	241	TAAACCTATGAAAGTGTGAGGATAGTGTTCGATATGAAATTAAGATTTCAACAAGTAG	300
Db	107780	CTTGAATTTTACAGTACTGATTTGCAATCTGACATGACGAATTTCTGTGCACTAATGT	1077211
OY	301	TAGCTGCTATGAAATTTAAGATTATTTATTTACACT--ATTATATAAATTTTAA	358
Db	107720	GGAAAGATTTCTTAGACTAAAGGACATCCAGAAATATTTCAACATCATTAAGTTTCA	1076611
OY	359	ACTATATACCTTAAATTTATTAAGGCTTTGAAATGGCGCAGCGCAGTAGAGCTCCGCT	418
Db	107660	TGACTTTCAATTTAAATACAGAAATTTAAAGAAATTTGGCGAGGTGAGGTCTCAGGCT	1076011
OY	419	GTAATCCCAACACTTTGGGAGGCGCAAGTGGCGGATCATCTGAGGCTCAGAGTTTAA	478
Db	107600	GTAATCCCAACACTTTGGGAGGCGCAGTGAAGATTCCTTGAGGCTCAGAGTTCAAA	1075411
OY	479	CCAGCTGGCCAAATGTTGAAACCTGTCTCTACTTAAACGCAAAATTTAGCCAGGTG	538
Db	107540	CCAGCTGGCCAAATGTTGAAACCTCATCTCTACTTAAATTAACAATTTAGCTGATG	1074811
OY	539	TGTTGGCATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGAGAAATTTGTTGA	598
Db	107480	TGTTGGCATGCACTGTGTAGTCCCACTTACTCAGGAGGCTGAGGACGAAATTTGTTGA	1074211
OY	599	CCTAGGAGGTGAGGTTGACATTAACCCGAG--ATGTCATGCACTGCACGCGCT--GGCA	654
Db	107420	CCGAGGAGGACAGGTTGACATGAGCCAGATCATGCCACTGCATCTTCAGCCTTGCGCAAC	1073611
OY	655	AGAGCAAGATTCATTAAGACAACAACCTTTGAAATTTGTGTAATGATTTACTTAT	714
Db	107360	AGAGCAAGCCTGTCTGAAAAATTTAAAAAATTTAATTTAAAAAATTTGGTGTATATTT	1073011
OY	715	CTTCACTTTAAGAAATTCAT	733
Db	107300	AGTAACTTTTAAAGTAAT	107282

RESULT 18	AC022445	AC022445	180907 bp	DNA	linear	PRI 18-DEC-2001
LOCUS	AC022445	AC022445	180907 bp	DNA	linear	PRI 18-DEC-2001
DEFINITION	AC022445	AC022445	180907 bp	DNA	linear	PRI 18-DEC-2001
ACCESSION	AC022445	AC022445	180907 bp	DNA	linear	PRI 18-DEC-2001
VERSION	AC022445.6	AC022445.6	180907 bp	DNA	linear	PRI 18-DEC-2001
KEYWORDS	HTG.	HTG.	180907 bp	DNA	linear	PRI 18-DEC-2001

Source	Organism	Reference Authors	Reference Title	Reference Journal	Reference Authors	Reference Title	Reference Journal	Comment
Source	Homo sapiens (human)	Homo sapiens						
Organism	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.							
Reference Authors	1 (bases 1 to 180807)	DOE Joint Genome Institute and Stanford Human Genome Center.						
Reference Title	Direct Submission							
Reference Journal	2 (bases 1 to 180807)	DOE Joint Genome Institute.						
Reference Authors	Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA							
Reference Title	3 (bases 1 to 180807)	DOE Joint Genome Institute and Stanford Human Genome Center.						
Reference Journal	Direct Submission							
Reference Authors	Submitted (28-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA							
Reference Title	4 (bases 1 to 180807)	DOE Joint Genome Institute and Stanford Human Genome Center.						
Reference Journal	Direct Submission							
Comment	Submitted (18-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA							
	On Dec 18, 2001 this sequence version replaced gi:15799584.							
	Draft Sequence Produced by DOE Joint Genome Institute							
	www.jgi.doe.gov							
	Finishing Completed at Stanford Human Genome Center							
	www.sbgc.stanford.edu							
	Quality: Phrap Quality >=40 99.7% of Sequence;							
	Estimated Total Number of Errors is 0.4.							
Features	Location/Qualifiers							
Source	1. 180807							
	/organism="Homo sapiens"							
	/mol_type="genomic DNA"							
	/db_xref="taxon:9606"							
	/chromosome="15"							
	/clone="RP11-489L13"							
Origin								
Query Match	11.2%; Score 224.2; DB 8; Length 180807;							
Best Local Similarity	60.4%; Pred. No. 1.5e-36;							
Matches	446; Conservative 0; Mismatches 278; Indels 15; Gaps 4;							
QY	1 TGATTCACGAGCCTTGAGCCTCCCAAGTGTCTGGAGTTACAGCGCTGAGCCACGAGCCTG 60							
DB	23342 TGATTCGCTGCTCAGCCTCTGGAAGTGTCTGGAGTTACAGCGCTGAGCCACGAGCCTG 23401							
QY	61 GTGCAATGCTTTATTAATTGGAAGACAACTGGGCTTAAATCTGTCCTTATTTGAC 120							
DB	23402 GCCATTAAGCTTTTTTAAACCAATCTG-----TTAGCCACATCTTAATTGTT 23452							
QY	121 AGACTTGTGATGAGTCAATGCCAATGCTGCCACTTACGACGGCTTAAATGACTTGA 180							
DB	23453 AAATATTAATCCATTAATCAACCAATAGACAGGGCAGAGAACTATATATCTGTA 23512							
QY	181 TCTCTCAGCCTGCTCTTTCGATATGTAAGGTGAATATATGAGCTTCAAGAGGA 240							
DB	23513 TTAATTTTGTGTGTTTTTAACCTTTTGTCTTCAATTTTAACTATTTTAAATAGAA 23572							
QY	241 TAAACCTATGAAAAGTGTGAGGATAGTGTGATATGTAATTAAGATTTTCAACAAGTAG 300							
DB	23573 CTGGAATTTTACAGTACGATGCTGATCTGACATGACGAATTTTCTGTGACATTAAGT 23632							
QY	301 TACCTGCTATGAAAGTTTAAGATTAATTATTAACACT--ATTAAATTAATTTTAAAA 358							
DB	23633 GGAAGATTTTCTTAGTACTTAAGAGCATCAAGAAATTTTCAACAATCATTAAGTTTCA 23692							
QY	359 ACTAATATCACTTAATTAATTAAGAGCTTTGAAATGGGCAAGCCAGTAGCTCTGCGCT 418							
DB	23693 TGACTTCAATTAATTAAGAAATTAAGAAATTTGGGCAAGTGAAGTGGCTCAGGCT 23752							
QY	419 GTATATCCAACTTTGGAGGCGCAAGTGGCGGATCACTGAGGTGACAGATTAAAGA 478							

Db	23753	GTATATCCAGCATCTTTGGAGGCCGAGCTAGAGATCACTTAGGTCAAGATTCAABA	23812
QY	479	CCAGCCTGGCCAACTGGTGAACCCCTGTCTTACTAAAAACGCAAAATTAGCCAGGTG	538
Db	23813	CCAGCCTGGCCAACTAGTGAACCTCATCTTAATAAAATTCAAAAATTAGCTGAGTG	23872
QY	539	TGTTGGCATGACCTGTATGTCTCCCACTACTCAGAGGTTGAGGAGAGAAATTCCTGAA	598
Db	23873	TGTTGGCATGACCTGTATGTCTCCCACTACTCAGAGGCTAGGACAGAGAAATTCCTGAA	23932
QY	599	CCTAGAGAGGTGGAGTTGACGTACCCGAG---ATGTCATGCATCCAGCTT-GGCAAC	654
Db	23933	CCGAGGAGGAGAGAGTTGACGTAGCCAGATATGTCATCTGACCTTCACTGGGCAAC	23992
QY	655	AGAGCAAGACTCCATAAAGACAAACAAAGCTTTGAAATGTGTAAATGAGTTGACTTAT	714
Db	23993	AGAGCAAGAGACCTCTCTGAAAAAATTAAATAAATTAATTAATAAATGGGTGTTATTT	24052
QY	715	CTTCATTTAGAAATTCAT 733	
Db	24053	AGTAACATTTTAAAGTAT 24071	
RESULT 19			
AC159033/c			
LOCUS	165467 bp	DNA	linear HTG 24-MAR-2005
DEFINITION	Pan troglodytes chromosome UNKNOWN clone CH251-339G2, WORKING DRAFT		
SEQUENCE	8 unordered pieces.		
AC159033			
AC159033.1	GI:61741063		
HTG, HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.			
Pan troglodytes (chimpanzee)			
Pan troglodytes			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
Hominiidae; Pan.			
1 (bases 1 to 165467)			
Wilson,R.K.			
The sequence of Pan troglodytes clone			
Unpublished			
2 (bases 1 to 165467)			
Wilson,R.K.			
Direct Submission			
Submitted (24-MAR-2005) Genetics, Genome Sequencing Center, 4444			
Forest Park Parkway, St. Louis, MO 63108, USA			
COMMENT			
----- Genome Center -----			
Center: Washington University Genome Sequencing Center			
Center code: WUGSC			
Web site:http://genome.wustl.edu			
Contact: submissions@wustl.edu			
----- Project Information -----			
Center project name: C_AB0339602			
----- Summary Statistics -----			
Sequencing vector: M13; 0%			
Sequencing vector: plasmid; 100%			
Chemistry: Dye-primer RT; 0% of reads			
Chemistry: Dye-terminator Big Dye; 100% of reads			
Assembly program: Phrap; version 0.990319			
Consensus quality: 160219 bases at least Q40			
Consensus quality: 161271 bases at least Q30			
Consensus quality: 161983 bases at least Q20			
----- NOTE: This is a 'working draft' sequence. It currently			
* consists of 8 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			
* This record will be updated with the finished sequence			
* as soon as it is available and the accession number will			
* be preserved.			

FEATURES	source	Location/Qualifiers
*	1	1371: contig of 1371 bp in length
*	1372	1471: gap of unknown length
*	1472	3143: contig of 1672 bp in length
*	3144	3243: gap of unknown length
*	3244	4584: contig of 1341 bp in length
*	4585	4684: gap of unknown length
*	4685	6388: contig of 1704 bp in length
*	6389	6488: gap of unknown length
*	6489	33801: contig of 2733 bp in length
*	33802	33901: gap of unknown length
*	33902	59513: contig of 25612 bp in length
*	59514	59613: gap of unknown length
*	59614	84765: contig of 25152 bp in length
*	84766	84865: gap of unknown length
*	84866	165467: contig of 80602 bp in length.
FEATURES	source	Location/Qualifiers
	1.	165467
		/organism="Pan troglodytes"
		/mol_type="genomic DNA"
		/db_xref="taxon:9598"
		/chromosome="UNKNOWN"
		/clone="CH251-33952"
		1..1371
misc_feature		/note="assembly_name:Contig10"
gap		1372..1471
		/estimated_length=unknown
misc_feature		1472..3143
gap		/note="assembly_name:Contig35"
		3144..3243
		/estimated_length=unknown
misc_feature		3244..4584
gap		/note="assembly_name:Contig36"
		4585..4684
		/estimated_length=unknown
misc_feature		4685..6388
gap		/note="assembly_name:Contig42"
		6389..6488
		/estimated_length=unknown
misc_feature		6489..33801
		/note="assembly_name:Contig43
		clone_end:SP6
		vector_side:left"
gap		33802..33901
		/estimated_length=unknown
misc_feature		33902..59513
		/note="assembly_name:Contig44"
gap		59514..59613
		/estimated_length=unknown
misc_feature		59614..84765
gap		/note="assembly_name:Contig45"
		84766..84865
		/estimated_length=unknown
misc_feature		84866..165467
		/note="assembly_name:Contig46
		clone_end:T7
		vector_side:left"
ORIGIN		
Query Match	11.2%;	Score 223.2; DB 14; Length 165467;
Best Local Similarity	85.1%;	Pred. No. 2.4e-36;
Matches 262;	Conservative 0;	Mismatches 43; Indels 3; Gaps 1;
QY	380	AAGAGCTTTGAAATGGCGCCAGGCGGACGTCCTCCGCTGTATCCCAACACTTTGGAG 439
DB	38467	AAAACTTTTGTCTTCGGCCAGGACAGTGGCTCATCGTGATATCCAGCACTTTGGAG 38408
QY	440	GCCAAAGGTGGCGGATACCTGAGGTCAAGGATTTAAGACCAAGCCCTGGCCAAACATGTGA 499
DB	38407	GTCAGGTGGGTGATACCTGAAGTCAAGATTTCAAGACCAAGCCCTGGCCAAACATGTGA 38348
QY	500	AACCTGTCTCTACTAATAAAACGAAAAATTTGCGCAGGTGTGTGTCATGCACCTGTGATC 559
DB	38347	AACCTGTCTCTACTAATAAAATACAAAAATTAGCCAGGCTTGGTGGCGGCGCCCTGTGATC 38288

Oy	560	CCAACTACTCAGAGAGTTGAGGGAGAGGAATTCGTGTAACCTTGAGAGAGTGGAGTGGAG	619
Db	38287	CCACTACTCAGAGAGCGCTGAGCAGAGAGAATTCGTTGAACCCAGAGGCGAGGCTTGCAG	38222
Oy	620	TAAACCCGAGAT--GTCACTGCACTCCAGCCTGGCAACAGAGCAAGATTCTAATAAGACA	676
Db	38227	TGACCCTCAGATCACGCCACTGCACTCCAGCCTGGAGACAGATGAAGCTCTGAAAAAAA	38168
Oy	677	ACAAAGC	684
Db	38167	AACAAAC	38160
RESULT 20			
LOCUS	AC079177		
DEFINITION	Homo sapiens Xp BAC Rpl1-147B14 (Roswell Park Cancer Institute		
ACCESSION	AC079177		
VERSION	AC079177.21	GI:13489128	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
	Hominidae; Homo.		
	1 (bases 1 to 155533)		
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,		
	Alsbrooks,S.L., Amaralungue,H.C., Are,J.T., Banks,T., Barbarella,J.,		
	Benton,J., Bimage,K., Blankenburg,K., Bonini,D., Bouck,J.,		
	Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bubay,C.,		
	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,		
	Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,		
	Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,		
	Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,		
	Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,		
	Denn,A.L., Ding,Y., Dinh,H.H., Doughaite,K.J., Draper,H.,		
	Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,		
	Elhaj,C., Becotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,		
	Foster,P., Frantz,P., Gabisi,A., Gao,U., Garcia,A., Garner,T.,		
	Gaston,N., Gill,R., Gorrell,J.H., Guevara,W., Gunnaratne,P., Hale,S.,		
	Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., HawesJA.,		
	Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,		
	Hollins,B., Homsi,P., Howard,S., Huber,J., Huliy,S., Hume,J.,		
	Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,		
	Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,		
	Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., LewisJ.L.C.,		
	Lewis,L., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W.,		
	Lozagedad,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,		
	Me,J., Mehevari,M., Mapua,P., Martin,R., Martindale,A.,		
	Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,		
	Mei,G., Metker,M., Miner,S., Moser,M., Neal,D., Newton,J., Nobhabat,K.,		
	Morgan,M., Morris,S., Moser,M., Neel,Z., Mitchell,T., Newson,N.,		
	Nuyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenko,S.,		
	Ogou,M., Okunnu,G., Oragune,N., Oviedo,R., Pace,A., Payton,B.,		
	Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pul,L.,		
	Quiles,M., Ken,Y., Rivers,M., Rojas,A., Rojwoodkan,I., Rolle,M.,		
	Raitz,S., Saevery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,		
	Stenson,I., Sodergren,E., Sonalik,T., Sparks,A., Stanley,H.,		
	Stone,H., Sutton,A., Svatek,A., Tabori,P., Tanerisa,A., Tanerisa,K.,		
	Tang,H., Tansie,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,		
	Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,		
	Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,		
	Washington,S., Williams,G., Williamson,A., Wiczek,R., Woodel,S.,		
	Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.		
	and Gibbs,R.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 155533)		
AUTHORS	Worley,K.C.		
JOURNAL	Submitted		
	Submitted (23-AUG-2000) Human Genome Sequencing Center, Department		

```

REFERENCE
AUTHORS
TITLE
JOURNAL
3 (bases 1 to 155533)
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
Worley,K.C.
Direct Submission
Submitted (30-MAR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 155533)
Worley,K.C.
Direct Submission
Submitted (07-APR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 155533)
Worley,K.C.
Direct Submission
Submitted (01-MAY-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 155533)
Worley,K.C.
Direct Submission
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 30, 2001 this sequence version replaced gi:13375595.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

COMMENT
JOURNAL
AUTHORS
TITLE
JOURNAL
CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) For Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT-----
----- Summary Statistics -----
Contig length: 155533
Phrap values in estimate: 154305
Average error rate (BCM-Phrap estimate): 2.34429e-06
Fraction of Phrap values less than 40 : 0.00316905
Number of consensus changing edits: 10

```

Summary Statistics	
Contig length:	15553
Phrap values in estimate:	154305
Average error rate (BCM-Phrap estimate):	2.34429e-06
Fraction of Phrap values less than 40 :	0.00316505
Number of consensus changing edits:	10

Number of N's in consensus :

0

Position	Consensus	Changing	Edits	Context
36096	gatacaacg	(n)	tgaccccccat	gatacaacg (c) tgaccccccat
36145	aaccacat	(g)	cccaatatc	aaccacat (c) cccaatatc
36460	atgccacg	(n)	cagaataatc	atgccacg (t) cagaataatc
45749	gaccttgg	(n)	nnnagcaatc	gaccttgg (t) tccgaataatc
45750	ccctctgg	(n)	nnnagcaatc	ccctctgg (t) cagaataatc
45751	cccttggm	(n)	nnagcaatc	cccttgg (c) cagaataatc
45752	tccttggm	(n)	nnagcaatc	tccttgg (c) cagaataatc
45753	atcaatct	(n)	nnagcaatc	atcaatct (n) gacagttctc
65973	tgatcagc	(n)	taagttctc	tgatcagc (c) taagttctc
153681	gltctcag	(n)	acatttggga	gltctcag (t) acatttggga

TITLE Direct Submission  
JOURNAL Submitted (10-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 106499)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 4 (bases 1 to 106499)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA (bases 1 to 106499)  
REFERENCE Wilson, R.K.  
TITLE Direct Submission  
JOURNAL Submitted (16-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
On Jul 3, 2001 this sequence version replaced gi:14165364.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@watson.wustl.edu  
----- Summary Statistics  
Center project name: H\_NH0096E08  
-----  
NOTICE:  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.  
-----  
MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu  
-----  
SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, B., Tateo, M., Catalanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org  
VECTOR: pBAC3.6  
-----  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-544H14, 2000 bp overlap; the clone sequenced to the right is RP11-33N4, 2000 bp overlap. Actual end of this clone is at base position 24907 of RP11-33N4.  
-----  
Data from AC080084 was used to finish RP11-96E8.  
Location/Qualifiers  
1..106499  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-96E8"  
/clone\_1kb="RPCI-11"  
118..1759  
/note="Cpg\_island (%GC=68.5, o/e=0.74, #CpGs=58) "  
-----  
FEATURES  
source  
misc\_feature

ORIGIN  
Query Match 11.1%; Score 222; DB 8; Length 106499;  
Best Local Similarity 61.7%; Pred. No. 4,6e-36;  
Matches 489; Conservative 0; Mismatches 290; Indels 13; Gaps 8;  
-----  
QY 2 GATCCACGACCTTGGCTCCCAAGTGTGGATTAACAGGCGTGAACCAACGCGCTGG 61  
DB 63875 GATTCCTCGCTCAGCTCCCGAGTAGCTGGATTACAGGACCTGCCACAGCCG 63816  
QY 62 TCGAATGCTTATTAATTTGAAGAGCAACATGGCTTTAAA---TCTGCTTTATTT 117  
DB 63815 -CTAATTTTGTATTTTGTAGAGACTCAGGTGACCCAGCGCTCAGCTCCCAAG 63757  
QY 118 GACAGACTTGTAGAGTGAATATCCCAATGCTGCCACTTACTGAACGCGCTTAAT 177  
DB 63756 TCGTAGATTACAGGATAGGCACACACCCGCCCTTGT-CAGGCTTTATGCTGCT 63698  
QY 178 TAGTCTCTCAGCTGTCTTCTCGCATATGTAAGTGAATATGATGCTTTCAGAG 237  
DB 63697 TTTTCTTGTCTATGATGATGTAT-AGCTCTAAAAATATTAATATTTATATTT 63639  
QY 238 GAATAACCTATGAAAAAGTGAAGATAGTGTGATATGAATTAAGATTTCACAAG 297  
DB 63638 TAAAGAAAGTTACTATTGTAACAGATTTTGAATATTTAGTCTCGAGTGAACCA 63579  
QY 298 TGTAGCTGCTATTGAAGATTTAAGATTATTTATTAACAATTTAATTAATTTTAA 357  
DB 63578 GAAATGTTTGTCCAACTCTTCAATTTCTGTTGATAGT-GATGTGCAATTTTACT 63520  
QY 358 AACTAATACATTAATTTAATTAAGGCTTTAAATGGGCGAGGCGTACTCTGCGC 417  
DB 63519 AATGAGAGTTAATGTTCTTATGTAAGAGACATCTAGGCTGGCGAGTGTCTATG 63460  
QY 418 TGTAAATCCCAACTTTGGAGGCCAAGTGGCGGATCACTGAGTCAAGATTAAAG 477  
DB 63459 TGTATTCAGACACTTTGGAGGCCGAGCGAGGATCACTGAGTCAAGATTGAG 63400  
QY 478 ACCAGCTGGCCAACTGTTGAACCTGTCTCTTACTTAAACGCAAAATTAAGCAG 537  
DB 63399 ACCAGCTGGCCAACTGTTGAATCTGTCTCTTACTTAAATTAAGCAGG 63340  
QY 538 GTGGTGGCATGACCTGTAGTCCCAACTCTGAGAGGTTGGAGAGGAATTTGCTTGA 597  
DB 63339 GTGGTGGCGGACCTGTATCCAGCTACTGAGAGGCTGAGGAGGAATCTTTGA 63280  
QY 598 ACTTAGAGAGTGTGAGTGTGACAGTAAACCGAGA---TGTACGTCACTCCAGCT- 653  
DB 63279 ACCGAGAGGCGAGAGTGTGACAGTGAATCTGTGCACTGCAATCCAGCTGGGCA 63220  
QY 654 CAGAGCAAGACTTCAT-AAAGCAACAAAGCTTTGAATTTGTAAATGAGTGTACCT 712  
DB 63219 CAGAGCAACACTTCATCAAAAAAAGGATCAATAGTACATATATGTAC 63160  
QY 713 ATCTTCAATTAAGAAATTCATCTTTGTCATTTATTTTACTGACATGAGCTTCGAG 772  
DB 63159 TCTTCATCATCATCTTTTGTCTTTTAACTTAATTAATTCAGACATTAATCTTT 63100  
QY 773 CAATTTTAATTT 784  
DB 63099 CTAATTTT 63088  
-----  
RESULT 22  
AC025750 151367 bp DNA linear PRI 30-APR-2005  
LOCUS Homo sapiens BAC clone RP11-804P20 from 2, complete sequence.  
DEFINITION AC025750  
ACCESSION AC025750.10 GI:18098549  
VERSION AC025750.10  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelostomi;



Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 151367)  
Cotton, M., Doebber, A. and Tomlinson, C.  
The sequence of Homo sapiens BAC clone RP11-804P20  
Unpublished (2001)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 151367)  
Waterston, R.H.  
Direct Submission  
Submitted (13-MAR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 151367)  
Waterston, R.  
Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

4 (bases 1 to 151367)  
Wilson, R.K.  
Direct Submission  
Submitted (30-APR-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT

On Jan 9, 2002 this sequence version replaced gi:13431263.

Center: Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu

Contact: submissions@watson.wustl.edu

Summary Statistics

Center project name: H\_NH0804P20

NOTICE:

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Fengen, E., Tateo, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.regen.com) or Pieter de Jong and coworkers at http://www.chori.org

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-299C5, 2000 bp overlap; the clone sequenced to the right is RP11-779M8. Actual start of this clone is at base position 157736 of RP11-299C5; actual end is at base position 151367 of RP11-804P20.

Data from AC074375 and AC006038 was used to finish this clone. AC025750. Polymorphisms have been identified between AC025750 and AC074375. The sequence from 40567 to 41115, from 40991 to 41115 are covered only by PCR products from clone DNA. There is an unresolved homopolymeric runs between 39229 and 39243. There is an unresolved region between 132476 and 132479.

FEATURES

Source 1..151367

ORIGIN

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-804P20"
/clone_11b="RPCT-11"
/complement(5917..57997)
/gene="KCNG3"
/complement(join(5917..8479,56737..57997))
/gene="KCNG3"
/complement(join(7834..8479,56737..57401))
/gene="KCNG3"
/note="Homo sapiens potassium voltage-gated channel,
subfamily G, member 3 (KCNG3), transcript variant 1,
mRNA.; H_NH0804P20.1
This gene was based on gi(27436991)"
/codon_start=1
/product="unknown"
/protein_id="AA24039.1"
/db_xref="gi:62988651"
/translation="MTFGRSGAASVNVNGARYSLRNLKDPPLRRVSRIRGCRSE
RDVLEVCDDYRENEYFEDRHSAPGFIILYRGHGRLRPAPMCISLYNEMLYWG
LEGAHLECCORRLDDMSPTVTVSADPEGLGRDEARPGAEAPSRVLEMRRT
FEPSITSLAQILASVSVYIVSMVVICASTLPDKRAAADNLSIDRSFSAQPR
EESGIIAICIGWFACTIVRIVSKNCEVKKPPLNITDLATPYTISLYMTFTG
AIFSAISQLBHGDLFETSNKDFISIPAAQWVLIISMTVGVGMVPTVGRILGV
CVSGIVLALPIPIYHSFVQCHELFKRSARSRSLSTFPLN"
56545..58926
/note="CpG island (%GC=71.6, o/e=0.94, #CpGs=297)"
132370..132990
/note="CpG island (%GC=77.0, o/e=0.87, #CpGs=67)"
132585..143100
/gene="tmp locus 6"
join(132585..132573,134337..134404,143007..143100)
/gene="tmp locus 6"
join(132646..132673,134337..134404,143007..143100)
/gene="tmp locus 6"
/note="Homo sapiens, similar to metacaspase associated 1,
clone MGC:4778 IMAGE:3542706, mRNA, complete cds.;
H_NH0804P20.2
This gene was based on gi(13278950)
Continues as H_NH0729M08.1"
/codon_start=1
/product="unknown"
/protein_id="AA24040.1"
/db_xref="gi:62988652"
/translation="MAAMNTRVGDYVYFENSSNPLYLRIBELNKTSQNVKAYVC
FYRRDISNLTLMADKHA"

Query Match 11.1%; Score 222; DB 8; Length 151367;
Best Local Similarity 60.6%; Pred. No. 4,3e-36;
Matches 417; Conservative 0; Mismatches 265; Indels 6; Gaps 3;

QY 1 TGATCCACGAGCTTGAGCTCCCAAGTGTGGATTAAAGCGGTGACGACGACGCTTG 60
DB 30890 TGATCCGCGCGCTCGGCTCCCAAGTGTGGATTAAAGCGGTGACGACGACGCTTG 30949

QY 61 GTGCAATGCTTATTAATTTAAAGACCAATGCGCTTAATCTGTCTTATTTTGC 120
DB 30950 GCCAAAACGATCTTTTAAACATTAATAATATTAATTAACACCTCCATGAAGATAA 31009

QY 121 AGACTTGATGAGTCAATATCCCAATGCTGCCCTTAAGAACGCGCTTAATGACTAG 180
DB 31010 TGTGAATAATTAGTGTAAGACCAATTTCCGAAAAAATGAATCAGACCAAAATTTTCA 31069

QY 181 TCTCTCTCAGCTGCTTTTGCATATGTAAGTGAATATGATGCTTTCAAGAGGAA 240
DB 31070 GTTTAGAAATTAACGAAAAACGACGCTTACATCTGTAAGAAAAAATTTGATAC 31129

QY 241 TAAACTATGAAAAAGTTGAGGATAGTGTGATATGAATTAAGATTTCACAACTAG 300

```

```

Db 31130 CTTGACATCTATTCACCAAAAGATGATACCGGCTCAGATGATTTTACAGCAAGTTC 31189
Oy 301 TAGCTGATTGAGAT--TTAAGATTATTATTAACAATTTAATTAATTTTAA 358
Db 31190 TATCATATACAGACATGAATTAATTCATTTTATTAATTAACACTGTCCTCCAAATTAATTA 31249
Oy 359 ACTAATACACTTAATTAATTAAGAGCTTTGAATGGGCGAGCGCAGTAGCTGCTGCT 418
Db 31250 AAGCATGACCAACTCATTTTAAAGATGAGTGAAGGCGAGTCCGAGCACTCAGCGCT 31309
Oy 419 GTAAATCCCAACACTTTGGAGGCGCAAGTGGCGGATCCTGAGTCAAGAGTTTAA 478
Db 31310 GTAAATCCCAACACTTTGGAGGCGCAAGTGGCGGATCCTGAGTCAAGAGTTTAA 31369
Oy 479 CCAGCTGCGCAACATGAGTGAAGACCTGCTCTACTAATAAAGCAAAATTTGCCAGTNG 538
Db 31370 CCAGCTGCGCAACATGAGTGAAGACCTGCTCTACTAATAAATGAGCAAGGCA 31429
Oy 539 TGGTGGCATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGGAATTTGCTTGA 598
Db 31430 TGGTGGCATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGGAATTTGCTTGA 31489
Oy 599 CTTAGAGGTGAGGTTGAGTGAACCCGAG--ATGTCACTGCACTCCAGCTT-GGCAAC 654
Db 31490 CCCAGAGAGGTGAGGTTGAGTGAACCTGATCATGCTGCACTCCAGCTGAGTGC 31549
Oy 655 AGAGCAAGACTCCATTAAGACACAA 682
Db 31550 AGAGTGAAGCTGTTCAAAAAA 31577

```

```

RESULT 23
AP003164
LOCUS 90289 bp DNA linear HTG 16-AUG-2001
DEFINITION Homo sapiens chromosome 1 clone dJ756A17 map 1p36.2-p36, ***
ACCESSION AP003164
VERSION AP003164.1 GI:15208266
KEYWORDS HTG; HTGS_PHASE1
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1
AUTHORS Machida,T., Ohira,M., Morohashi,A., Mihara,M., Furuta,S., Seeda,E.
and Nakagawara,A.
TITLE Homo sapiens 90,289 genomic DNA of 1p36.2-p36
JOURNAL Published Only in Database (2001)
AUTHORS 2 (bases 1 to 90289)
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2001) Akira Nakagawara, Chiba Cancer Center
AUTHORS Research Institute, Division of Biochemistry; 666-2 Nihona,
Chunoh-ku, Chiba, Chiba 260-8717, Japan
(E-mail:akiramakchiba-cc.pref.chiba.jp, Tel:81-43-264-5431,
Fax:81-43-265-4459)
COMMENT Additional author information
Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A.,
Yada,T., Toroki,Y., Watanabe,H. and Sakaki,Y.
The Institute of Physical and Chemical Research (RIKEN) Genomic
Sciences Center (GSC)
1-7-22 Saitoh-chou,Tsukumi-ku,Yokohama,Kanagawa, JAPAN zip:
230-0045
phone: 81-45-503-9111, fax: 81-45-503-9170
e-mail: hattori@gsc.riken.go.jp

```

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence

as soon as it is available and the accession number will be preserved

```

1 13927 contig of 13927 bp in length
14028 22029 contig of 8002 bp in length
22130 56617 contig of 34488 bp in length
56718 66346 contig of 9629 bp in length
66447 76543 contig of 10097 bp in length
76643 90289 contig of 13646 bp in length

```

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 13927 contig of 13927 bp in length
14028 22029 contig of 8002 bp in length
22130 56617 contig of 34488 bp in length
56718 66346 contig of 9629 bp in length
66447 76543 contig of 10097 bp in length
76643 90289 contig of 13646 bp in length.

```

#### FEATURES

```

source
1..90289
location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.2-p36"
/clone="dJ756A17"

```

#### ORIGIN

```

Query Match 11.1%; Score 221.2; DB 14; Length 90289;
Best Local Similarity 83.8%; Pred. No. 7e-36;
Matches 263; Conservative 0; Mismatches 48; Indels 3; Gaps 1;

Oy 388 TGAATGGCGCAGGCGAGTGCCTGCTGATATCCCAACACTTTGGAGGCGCAAGT 447
Db 20511 TGAAGTCGCGCAGGCGAGTGCCTGCTGATATCCCAACACTTTGGAGGCGCAAGT 20570
Oy 448 GGGCGGATCACCTGAGTCAAGAGTTTAAAGACCAAGCTGGCCCAACATGAGTAA 507
Db 20571 GGGCGGATCACCTGAGTCAAGAGTTTAAAGACCAAGCTGGCCCAACATGAGTAA 20630
Oy 508 CTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 567
Db 20631 CTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 20690
Oy 568 TCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 627
Db 20691 TCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20750
Oy 628 GAT---GTCACTGCACTTCAGCTGGCAACAGACCAAGCTTCATTAAGACCAAAAGC 684
Db 20751 GATGGCGGCACTGCACTTCAGCTGGAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAG 20810

```

Oy 685 TTGAAATGGTGA 698  
|||||  
Db 20811 AAATTAATTCGAA 20824

RESULT 24  
AX039602 121162 bp DNA linear PAT 18-NOV-2000  
LOCUS Sequence 1 from Patent WO0063375.  
DEFINITION AX039602  
ACCESSION AX039602  
VERSION AX039602.1 GI:11229631  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1  
REFERENCE  
AUTHORS Bougueterec, L., Dufaure-Gare, I., and Grel, P.  
TITLE Dna encoding a kinesin-like protein (hklp) comprising diallelic markers  
JOURNAL Patent: WO 0063375-A 1 26-OCT-2000;  
GENSET (FR)  
FEATURES  
source  
1..121162  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
4844..4920  
/note="exon"  
7041..7060  
/note="12-809.pu"  
7058..7237  
/note="exon"  
7140..7158  
/note="12-809-119.mis"  
7147..7171  
/bound\_moiety="12-809-119.probe"  
variation  
7159  
/note="12-809-119 : polymorphic base G or C"  
primer\_bind  
7160..7178  
/note="12-809-119.mis complement"  
7545..7565  
primer\_bind  
/note="12-809.rp complement"  
10469..10534  
/note="exon"  
15940..16118  
/note="exon"  
16004..16024  
/note="12-805.rp"  
16350..16368  
primer\_bind  
/note="12-805-115.mis"  
16357..16381  
/bound\_moiety="12-805-115.probe"  
16369  
variation  
/note="12-805-115 : polymorphic base A or G"  
16370..16388  
/note="12-805-115.mis complement"  
16464..16483  
primer\_bind  
/note="12-805.pu complement"  
16712..16823  
exon  
/note="exon"  
20054..20131  
exon  
/note="exon"  
20793..20858  
exon  
/note="exon"  
21558..21575  
exon  
/note="exon"  
23971..24046  
exon  
/note="exon"  
24864..24942  
exon  
/note="exon"

primer\_bind 26232..26252  
/note="12-790.rp"  
primer\_bind 26291..26309  
/note="12-790-396.mis"  
misc\_binding 26298..26322  
/bound\_moiety="12-790-396.probe"  
variation 26310  
/note="12-790-396 : polymorphic base A or G"  
26311..26329  
/note="12-790-396.mis complement"  
26529..26671  
exon  
/note="exon"  
26855..26705  
/note="12-790.pu complement"  
27640..27681  
exon  
/note="exon"  
30863..31074  
exon  
/note="exon"  
30902..30922  
primer\_bind  
/note="12-791.pu"  
31093..31111  
primer\_bind  
/note="12-791-211.mis"  
31100..31124  
misc\_binding  
/bound\_moiety="12-791-211.probe"  
31112  
variation  
/note="12-791-211 : polymorphic base A or G"  
31113..31131  
primer\_bind  
/note="12-791-211.mis complement"  
31282..31301  
primer\_bind  
/note="12-791.rp complement"  
33476..33496  
primer\_bind  
/note="12-803.rp"  
33789..33807  
primer\_bind  
/note="12-803-125.mis"  
33796..33820  
misc\_binding  
/bound\_moiety="12-803-125.probe"  
33808  
variation  
/note="12-803-125 : polymorphic base A or T"  
33809..33827  
primer\_bind  
/note="12-803-125.mis complement"  
33912..33932  
primer\_bind  
/note="12-803.pu complement"  
33934..33953  
primer\_bind  
/note="99-33040.pu"  
34236..34254  
primer\_bind  
/note="99-33040-321.mis"  
34243..34267  
misc\_binding  
/bound\_moiety="99-33040-321.probe"  
34255  
variation  
/note="99-33040-321 : polymorphic base C or T"  
34256..34274  
primer\_bind  
/note="99-33040-321.mis complement"  
34364..34383  
primer\_bind  
/note="99-33040.rp complement"  
34918..34938  
primer\_bind  
/note="12-810.rp"  
35274..35292  
primer\_bind  
/note="12-810-77.mis"  
35281..35305  
misc\_binding  
/bound\_moiety="12-810-77.probe"  
35293  
variation  
/note="12-810-77 : polymorphic base A or G"  
35294..35312  
primer\_bind  
/note="12-810-77.mis complement"  
35350..35369  
primer\_bind  
/note="12-810.pu complement"  
39625..39704  
exon  
/note="exon"  
40590..40665  
exon  
/note="exon"  
43630..43709  
exon  
/note="exon"  
44204..44310  
exon

```
/note="exon"
45126..45209
/note="exon"
45441..45621
exon
/note="exon"
45465..45485
/note="12-787-787"
45718..45790
exon
/note="exon"
45873..45891
primer_bind
/note="12-787-103.mis"
45880..45904
misc_binding
/bound_moiecy="12-787-103.probe"
45892
variation
/note="12-787-103 : polymorphic base A or G"
45893..45911
primer_bind
/note="12-787-103.mis complement"
45974..45993
primer_bind
/note="12-787.pu complement"
54361..54381
primer_bind
/note="12-793.rp"
54478..54496
primer_bind
/note="12-793-383.mis"
54485..54509
misc_binding
/bound_moiecy="12-793-383.probe"
54497
variation
/note="12-793-383 : polymorphic base G or T"
54498..54516
primer_bind
/note="12-793-383.mis complement"
54861..54879
primer_bind
/note="12-793.pu complement"
56365..56385
primer_bind
/note="12-792.rp"
56563..56581
primer_bind
/note="12-792-233.mis"
56570..56594
misc_binding
/bound_moiecy="12-792-233.probe"
56582
variation
/note="12-792-233 : polymorphic base A or G"
56583..56601
primer_bind
/note="12-792-233.mis complement"
56793..56813
primer_bind
/note="12-792.pu complement"
60111..60130
primer_bind
/note="99-41009.rp"
60317..60335
primer_bind
/note="99-41009-244.mis"
60324..60348
misc_binding
/bound_moiecy="99-41009-244.probe"
60336
variation
/note="99-41009-244 : polymorphic base A or G"
60337..60355
primer_bind
/note="99-41009-244.mis complement"
60450..60468
primer_bind
/note="99-41009-111.mis"
60457..60481
misc_binding
/bound_moiecy="99-41009-111.probe"
60469
variation
/note="99-41009-111 : polymorphic base C or T"
60470..60488
primer_bind
/note="99-41009-111.mis complement"
60562..60580
primer_bind

Query Match 11.1%; Score 221.2; DB 6; Length 121162;
Best Local Similarity 83.8%; Pred.No. 6,6e-36;
Matches 263; Conservative 0; Mismatches 48; Indels 3; Gaps 1;
```

```
Db 25395 GGGCGGATCAGTTCAGTTCAGGAGTTTGAAGACCAAGCTGGCCAAATGATGAACCCCAT 25454
Qy 508 CTTCTACTAAAAAGCAAAATTTAGCCAGGTGTGTGGTCATTCAGCTGTAGTCCCACTC 567
Db 25455 CTTCTACTAAAAATACAAAATATGATGTCGGCGTGTGCATTCAGCTGTATCTTACTC 25514
Qy 568 TCAGAGAGTTGAGGAGAGAAATTCCTTAACCTAGAGAGTTGAGGTTTCAGTAACCCGA 627
Db 25515 TCAGAGAGCTGAGGAGAGAAATTCCTTAACCTAGAGAGGCGGAGTTTCAGTAGACCGA 25574
Qy 628 GAT--GTCAGTGCAGTCCAGCTTGCAACAGAGCAAGATTCATTAAGACAAACAAAGC 684
Db 25575 GATGGCGCAGTGCAGTCCAGCTTGCAACAGAGCAAGATTCATTAAGACAAACAAAGC 25634
Qy 685 TTTGAAATTTGTGA 698
Db 25635 AATAAATTTCTGAA 25648
```

```
RESULT 25
LOCUS AL358013 171612 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-420G9 on chromosome 1. Contains the 5' end of the KIF1B gene for kinesin family member 1B and a CpG island, complete sequence.
ACCESSION AL358013
VERSION AL358013.19 GI:17017702
KEYWORDS HTG; KIF1B.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 171612)
```

```
REFERENCE
AUTHORS Dunn,M.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
```

## COMMENT

On Nov 20, 2001 this sequence version replaced gi:16973843. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-420G9 is from the library RPEC11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [vegas@sanger.ac.uk](mailto:vegas@sanger.ac.uk)

-----  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

```
source 1..171612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-420G9"
```



111539. .111719,111816. .111888,134685. .134778,  
136351. .136499,138526. .138704,139400. .139537,

Query Match 11.1%; Score 221.2; DB 8; Length 171612;  
Best Local Similarity 83.8%; Pred. No. 6; 2e-36;  
Matches 263; Conservative 0; Mismatches 48; Indels 3; Gaps 1;

QY 388 TGAATAGGGCCAGCCAGTAGCTCTGCTTAATCCCAACATTGGAGGCCAAGT 447  
DB 91434 TGAAGTCGGCCAGGCCAGTGGCCCCCACTGTATCCAGCACTTTGGGTGGCCAGGT 91493  
QY 448 GGGCGGATCACTGAGGTCAAGATTGAGCAAGCCAGTCCCAACATGTAACCCCTGT 507  
DB 91494 GGGCGGATCACTGAGGTCAAGATTGAGCAAGCCAGTCCCAACATGTAACCCCAT 91553  
QY 508 CTCTACTTAAAGGCAAAATTGAGCCAGTGTGTGTGATGACCTGTATCCCAATAC 567  
DB 91554 CTCTACTTAAAGGCAAAATTGAGCCAGTGTGTGTGATGACCTGTATCCCAATAC 91613  
QY 568 TCAGAGAGTTGAGGAGAGAAATTGCTTGAACCTAGAGGTGAGGTGACGTAACCCGA 627  
DB 91614 TCAGAGAGTTGAGGAGAGAAATTGCTTGAACCTAGAGGTGAGGTGACGTAACCCGA 91673  
QY 628 GAT--GTCACTGCACTCCAGCTGCGCAACAGCAAGACTCCATAAGCAACAAAGC 684  
DB 91674 GATGCGGCACCTGCACTCCAGCTGCGCAACAGCAAGACTCCATAAGCAACAAAGC 91733  
QY 685 TTTGAATTTGTCTA 698  
DB 91734 AATAAATTCTGAA 91747

RESULT 26  
AC137058/c 185623 bp DNA linear HTG 21-FEB-2003  
LOCUS Papio anubis clone RP41-126M5, WORKING DRAFT SEQUENCE, 10 ordered  
DEFINITION pieces.

AC137058 GI:28460766  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE Papio anubis (olive baboon)  
ORGANISM Papio anubis

REFERENCE  
1 (bases 1 to 185623)  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Cercopithecoidea; Cercopithecoidea; Papio.

REFERENCE  
AUTHORS Akher, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Cariaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,  
Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, B.,  
Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, O.L., Maduro, V.B.,  
Marquies, B.H., Masello, C., Maskeri, B., McDowell, J.,  
Pagittigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,  
Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,  
Stattrop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,  
Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.  
NISC Comparative Sequencing Initiative

TITLE NISC Comparative Sequencing Initiative  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 185623)  
AUTHORS Green, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (15-NOV-2002) NIH Intramural Sequencing Center, 8717  
Groveomont Circle, Gaithersburg, MD 20877, USA  
3 (bases 1 to 185623)  
REFERENCE Green, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (21-FEB-2003) NIH Intramural Sequencing Center, 8717  
Groveomont Circle, Gaithersburg, MD 20877, USA  
On Feb 21, 2003 this sequence version replaced gi:27476124.

COMMENT  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov

Contact: nisc.zoo@nhgri.nih.gov  
Project Information  
Center project name: dtj  
Center clone name: 126M05

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

#### Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 183092 bases at least Q40  
Consensus quality: 183851 bases at least Q30  
Consensus quality: 184434 bases at least Q20  
Insert size: 16000; agarose-fp  
Insert size: 184723; sum-of-contigs  
Quality coverage: 14.13x in Q20 bases; agarose-fp  
Quality coverage: 12.24x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 9984: contig of 9984 bp in length  
\* 9985 10084: gap of unknown length  
\* 10085 49805: contig of 39721 bp in length  
\* 49806 49905: gap of unknown length  
\* 49906 66936: contig of 17031 bp in length  
\* 66937 67037: gap of unknown length  
\* 67037 83604: contig of 16568 bp in length  
\* 83605 83704: gap of unknown length  
\* 83705 108246: contig of 24542 bp in length  
\* 108247 112021: gap of unknown length  
\* 112022 112121: gap of unknown length  
\* 112122 134817: contig of 22696 bp in length  
\* 134818 134917: gap of unknown length  
\* 134918 147290: contig of 12373 bp in length  
\* 147291 147390: gap of unknown length  
\* 147391 185337: contig of 37947 bp in length  
\* 185338 185437: gap of unknown length  
\* 185438 185623: contig of 186 bp in length.

#### Location/Qualifiers

1. 185623  
/organism="Papio anubis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9555"  
/clone="RP41-126M5"  
/clone\_1lb="RP41"  
1. 15263  
/note="clone overlaps with GenBank Accession Number  
AC130273 clone RP41-30706 (center project name dsu)"  
1. 9984  
/note="assembly\_fragment  
clone\_end:77  
vector\_side:left"  
9985..10084  
/estimated\_length=unknown  
gap

```

misc_feature      10085..49805
                    /note="assembly_fragment"
gap               49805..49905
                    /estimated_length=unknown
misc_feature      49905..66936
                    /note="assembly_fragment"
gap               66936..67036
                    /estimated_length=unknown
misc_feature      67036..83604
                    /note="assembly_fragment"
gap               83604..83704
                    /estimated_length=unknown
misc_feature      83704..108246
                    /note="assembly_fragment"
gap               108246..108346
                    /estimated_length=unknown
misc_feature      108346..112021
                    /note="assembly_fragment"
gap               112021..112121
                    /estimated_length=unknown
misc_feature      112121..134817
                    /note="assembly_fragment"
gap               134817..134917
                    /estimated_length=unknown
misc_feature      134917..147290
                    /note="assembly_fragment"
gap               147290..147390
                    /estimated_length=unknown
misc_feature      147390..185337
                    /note="assembly_fragment"
gap               185337..185437
                    /estimated_length=unknown
misc_feature      185437..185623
                    /note="assembly_fragment"
                    clone_end:Spe
                    vector_side:right"

ORIGIN
Query Match      11.0%; Score 221; DB 14; Length 185623;
Best Local Similarity 62.7%; Pred. NO. 6.7e-36;
Matches 432; Conservative 0; Mismatches 240; Indels 17; Gaps 5;

QY      1 TGATCCACGACCTTGCGCTCCCAAGTGTGGAGATTACAGCGGTGACCCACGCGCTG 60
      36364 TGATCCGCGCGTCTCGGCTCCCAAGTGTGGAGATTACAGCGGTGACCCACGCGCG 36305
      61 GTGGAATGCTTTATTTATTTGAAGAGACAACATGGCGCTTAANTCTGCTTTATTTGAC 120
      36304 GCCCTATTTATTTATTTTAAAAAATACAC-----CAACTACACATTTTGTGTGAG 36253
      121 AGACTTTGATGAGTCAATCCCAATGCTGCCACTTACGACGCGCTTAATGACTAG 180
      36252 GTATTTGTTACCTTTCTATCTCATACTTACACAGACTTGTGTTTCTTTCTCTTTA 36193
      181 TCTCTCAGCTGCTTTCTGCAATATGTAAGTGAATTAAGTCTTCAAGAGAA 240
      36192 TCTCTTGTATTTGAATTTCTTCCCAATGGGTTTCAATTTGGTTCTCAAGGCTC 36133
      241 TAAACCTATGAAAAAGTTGAGTACTGTTGATATGAATAAGATTTCAACAGTAG 300
      36132 TGAAGATTATATGCTCGAATAATATTTATCCCTTACATTTGAATTTATTTGCGAG 36073
      301 TAGCTGCTATTTGAAGATTATTTATTTCAACATTTAATAATTTTAAAAAC 360
      36072 GATATACATTTCTAAATTTAAAGATTTTCTTGTGCTTTAAACTTACCCCACTGTT 36013
      361 TAAATACCTTAAAT---ATTAAAGACTTTGAATGGGCGAGCGAGTAGCTCTGCC 417
      36012 TCTTGCAATTATGATCTGCTGCAAGAAATCTTAAGCGCGCAAGGTCCGGTGTTCAGCC 35953
      418 TGTATCCCAACACTTTGGAGGCGCAAGGTGGCGGATCACTGAGGTCAAGAGTTTAA 477
      35952 TGTATCCCAACTTTGGAGGCGCAAGGTGGCGGATCTTGAAGGTCAAGAGTTTCAAG 35893

```

```

QY      478 ACCAGCTGGCCCAACATGTTGAAACCTGCTCTTACTTAAAAACGCAAAATTTAGCCAGT 537
      35892 ACCAGCTGGCCCAACATGTTGAAACCTGCTCTTACTTAAAAA--TAAAAATAGCCGCGC 35835
QY      538 GTGTGGATGACCTGTATGATCCCAATCTTCAGAGGTTGAGGAGAGAAATTTCTTGA 597
      35834 GTGTGGATGACCTGTATGATCCCAATCTTCAGAGGTTGAGGAGAGAAATTTCTTGA 35775
QY      598 ACCTAGAGGTGAGGTTGACAGTAACCCAGAT---GTCACTGCACTCCAGCT--GGCA 653
      35774 ACCAGAGGAGAGAGGCTGTATGAGCCAAATATACGCCACCTGCACTCCAGCTGGCA 35715
QY      654 CAGACCAAGCTCCATTAAGACCAACAAA 682
      35714 CAGACTAGACTCCGCTTCAGAAAAAAA 35686
      Db

RESULT 27
AC020603
LOCUS      AC020603      202317 bp      DNA      linear      PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-534K13 from 8, complete sequence.
ACCESSION AC020603
VERSION   AC020603.4   GI:9454643
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homnidae; Homo.
REFERENCE 1 (bases 1 to 202317)
           Sulston,J.E. and Waterston,R.
           Toward a complete human genome sequence
           Genome Res. 8 (11), 1097-1108 (1998)
           9847074
REFERENCE 2 (bases 1 to 202317)
           Kruchowski,S., Stoneking,T., Gregory,S. and Markovic,C.
           The sequence of Homo sapiens BAC clone RP11-534K13
           Unpublished
           3 (bases 1 to 202317)
REFERENCE 3 (bases 1 to 202317)
           Waterston,R.H.
           Direct Submission
           Submitted (05-JAN-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
           5 (bases 1 to 202317)
REFERENCE 4 (bases 1 to 202317)
           Waterston,R.H.
           Direct Submission
           Submitted (08-NOV-2000) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           6 (bases 1 to 202317)
REFERENCE 5 (bases 1 to 202317)
           Waterston,R.
           Direct Submission
           Submitted (09-MAY-2001) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Jul 26, 2000 this sequence version replaced gi:7631048.
           ----- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: sapiens@wustl.wustl.edu
           ----- Summary Statistics
           Center project name: H_NH0534K13
           -----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping

```





/rpl\_family="LI"

Query Match 11.0%; Score 221; DB 8; Length 202317;  
Best Local Similarity 59.3%; Pred. No. 6, 6e-36;  
Matches 432; Conservative 0; Mismatches 290; Indels 7; Gaps 3;

1 TGATCCACAGCCTGGGCTCCCAAGTCTGGGATTAACAGGCGTAGCCAGCCCTG 60  
Db 86875 TGATCCACAGCCTGGGCTCCCAAGTCTGGGATTAACAGGCGTAGCCAGCCCA 86934

61 GTGCAATGCTTTATTAATTAATTAAGACAAATGGGCTTAAATCTGCTTAAATTTGAC 120  
Db 86935 GCCCGATGATTTTACTGCTCTCTCTCTCCCTAGTCTTTCTGCTTTTCCATCCGC 86994

121 AGACTTTAGTAGAGTCAATCCCAATGCTGCCACTTACGAACGGCTTAATGACTTNG 180  
Db 86995 TTTTCTGCTGCTCCCTCTTCAATTTCTGTTCTCTTCTCTCTCTCTTAAATTAATTT 87054

181 TCTCTCTGAGCTGTCTTTCTGCTATGTAAGTGGAATATGATGGCTTTCAAGAGAA 240  
Db 87055 GAGCATTTCAAAATATTTTAAAGCTGAGATGTTGAATGTTTATAGCTTGTCTCT 87114

241 TAAACCTATGAAAAGTGTGAGGATGTTGATATGAAATAGCAATTTCAACAAATG 300  
Db 87115 TAAATATATTAATAGTATGATTTATTTATTTCTCTATTTCTCCCAAGCAAAAGAAA 87174

301 TAGCTGCTATGAAATTTAAGGTTATTTATTAACAATTTAATTAATTAATTTAAAC 360  
Db 87175 TCTCCACAGTGGCAATTAATTAATTAAGATTTTAAATAATTTTGTGTTTACA 87234

361 TATATACATTAATTAATTAAGAGCTTTGAAT--GGGCGAGCGGAGTCTGCTGCT 418  
Db 87235 TTAATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 87294

419 GTATCCCAACACTTTTGGAGGCGCAAGTGGGCGGATCACTGAGTCAAGAGCTTTAAG 478  
Db 87295 GTATCCCAACACTTTTGGAGGCGTGAAGCGTGAAGTCACTGAGTCAAGAGCTTTAAG 87354

479 CCAGCTGGGCAACATGTTGAAACCTGTTCTTACTTAAATTAATTAATTAATTAATTA 538  
Db 87355 CCAGCTGGGCAACATGTTGAAACCTGTTCTTACTTAAATTAATTAATTAATTA 87414

539 TGTGTCATGACCTGTCATGTCCTCACTCACTCACTCACTCACTCACTCACTCACTCA 598  
Db 87415 TGTGTCATGACCTGTCATGTCCTCACTCACTCACTCACTCACTCACTCACTCACTCA 87474

599 CCTAGAGGTGAGGTTGACGTAACCCGAGA--TGTCACTGCACTCCAGCTTGG--CAA 653  
Db 87475 CCTAGAGGTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAG 87534

654 CAGAGCAAGCTCTCATTAAGACAAACAAAGCTTTGAATTTGTGAATTTGATGTTACTTA 713  
Db 87535 AAGAGCAAGCTCTCATTAAGACAAACAAAGCTTTGAATTTGTGAATTTGATGTTACTTA 87594

714 TCTTCAATTT 722  
Db 87595 ATTTCAATTT 87603

RESULT 28  
AC036222 174965 bp DNA linear PRI 19-OCT-2002  
DEFINITION Homo sapiens chromosome 17, clone RP11-433M22, complete sequence.  
AC036222  
VERSION AC036222.23 GI:24137556  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 174965)  
AUTHORS Birren,B., Nusbaum,C. and Lander,B.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Homo sapiens chromosome 17, clone RP11-433M22  
Unpublished  
2 (bases 1 to 174965)  
Birren,B., Linton,L., Nusbaum,C., Lander,B., Abraham,H., Allen,N.,  
Anderson,S., Baldini,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Boukhalter,B., Brown,A., Burnett,G.,  
Campiano,A., Castle,A., Choquet,Y., Colangelo,M., Collins,S.,  
Collins,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karatas,A.,  
Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKenna,K., McPheters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,C., Morrow,J.,  
Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,D., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 174965)  
Birren,B., Nusbaum,C., Lander,B., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,  
Camarda,J., Chang,J., Chazaro,B., Choquet,Y., Collamore,A.,  
Cooke,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,  
Mathews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,  
Mlenka,C., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,  
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnupack,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Talamas,J., Tefaye,S., Theodore,J., Topham,K.,  
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (15-OCT-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 174965)  
Birren,B., Nusbaum,C., Lander,B., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,  
Camarda,J., Chang,J., Chazaro,B., Choquet,Y., Collamore,A.,  
Cooke,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,  
Mathews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,  
Mlenka,C., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,  
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnupack,R.,  
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Talamas,J., Tefaye,S., Theodore,J., Topham,K.,  
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (19-OCT-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Oct 19, 2002 this sequence version replaced gi:23957647.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L9421  
Center clone name: 433\_M22  
-----

Only the last 175.0 kilobases of this clone are being submitted.  
The remainder overlaps accession number AC103702 [WICGR project  
L21590].

## FEATURES

source  
Location/Qualifiers  
1..174965  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-433M22"  
8..136  
/clone\_lib="RP11-433M22"  
repeat\_region  
/rpt\_family="L2"  
complement(436..549)  
/rpt\_family="MIR"  
repeat\_region  
570..859  
/rpt\_family="AluSc"  
repeat\_region  
860..888  
/rpt\_family="AT-rich"  
complement(1035..1254)  
/rpt\_family="L1MC4"  
repeat\_region  
complement(1299..1970)  
/rpt\_family="L1MB2"  
repeat\_region  
complement(1971..2266)  
/rpt\_family="AluSc"  
repeat\_region  
2267..2325  
/rpt\_family="TTG)n"  
repeat\_region  
complement(2328..2608)  
/rpt\_family="AluSc"  
repeat\_region  
complement(2609..2932)  
/rpt\_family="L1MB2"  
repeat\_region  
complement(2943..3230)  
/rpt\_family="AluSc"  
repeat\_region  
3287..3560  
/rpt\_family="L2"  
complement(3704..3857)  
/rpt\_family="MIR"  
repeat\_region  
3858..4159  
/rpt\_family="AluSp"  
complement(4160..4205)  
/rpt\_family="MIR"  
repeat\_region  
4206..4571  
/rpt\_family="L3"  
repeat\_region  
4677..4965  
/rpt\_family="AluJo"  
complement(6458..6602)  
/rpt\_family="MIR"  
repeat\_region  
6675..6778  
/rpt\_family="L1MC5"  
complement(6817..6859)  
/rpt\_family="L3"  
repeat\_region  
complement(6910..7127)  
/rpt\_family="MIR"  
complement(7141..7437)  
/rpt\_family="AluSc"  
complement(8922..9145)  
/rpt\_family="MIR"  
9466..9595  
/rpt\_family="FLAM\_C"  
10532..10633

/rpt\_family="MIR"  
12389..12450  
/rpt\_family="MERSA"  
repeat\_region  
12451..12761  
/rpt\_family="AluJo"  
repeat\_region  
12762..12860  
/rpt\_family="MERSA"  
repeat\_region  
complement(12869..12914)  
/rpt\_family="MIR"  
repeat\_region  
complement(12915..13156)  
/rpt\_family="AluSc"  
complement(13157..13310)  
/rpt\_family="MIR"  
repeat\_region  
13546..13604  
/rpt\_family="L2"  
repeat\_region  
14682..14725  
/rpt\_family="TATMA)n"  
repeat\_region  
14929..15043  
/rpt\_family="AluJo"  
repeat\_region  
15044..15074  
/rpt\_family="AT-rich"  
repeat\_region  
15463..15484  
/rpt\_family="L2"  
complement(15485..15787)  
/rpt\_family="AluY"  
repeat\_region  
15788..15869  
/rpt\_family="L2"  
complement(16233..16368)  
/rpt\_family="L1MA10"  
repeat\_region  
complement(16384..16538)  
/rpt\_family="MIR3"  
repeat\_region  
16644..16697  
/rpt\_family="CT-rich"  
complement(16771..17063)  
/rpt\_family="AluSc"  
repeat\_region  
complement(17066..17263)  
/rpt\_family="MIR"  
repeat\_region  
17267..17293  
/rpt\_family="TTTG)n"  
repeat\_region  
17425..17562  
/rpt\_family="L2"  
complement(18106..18304)  
repeat\_region

Query Match 11.0%; Score 220.6; DB 8; Length 174965;  
Best Local Similarity 61.2%; Pred. No. 8.2e-36;  
Matches 430; Conservative 0; Mismatches 259; Indels 14; Gaps 4;

QY 1 TGATCCACGACCTTGCGCTCCCAAGTGTGGATTACAGGCGTGAGCCACGACGCTG 60  
DB 68046 TGATCCACCGCGCTTACGCTTCCCAAGTGTGGATTACAGGCGTGAGCTGCGCCCG 68105  
QY 61 GTGCAATGCTTATTTATTTGAAGACACACATGGCGCTTAAATCTGCTTATTTGAC 120  
DB 68106 GCTCGTACTAGTGTATTTTATTTAGGGAAATTCCTCATTTAGAAATATCGTT -GA 68163  
QY 121 AGACTTTGATGAGTCACAAATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTT 180  
DB 68164 AGTTTGTAGGGAAGGAGCGGTGATGACTGTAACTCTCAAAATTCAGAAAAAATGT 68223  
QY 181 TCTCTCAGCTGCTCTTCTGATGATGTAAGGTGAATTAATGATGCTTCAAGAGGAA 240  
DB 68224 AAATTAATCTAGTACTAATTTACTTTCTTAATTAATCAATTTATTTATTAAGAAAAATA 68283  
QY 241 TAAACCTATGAAGAAGTGTGAGATAGTGTGATGATGAATTAAGATTTCAACAAGTAG 300  
DB 68284 AGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 68343  
QY 301 TAGCTGCTATTAAGATTTAAGATTTATTTATTAACAATA-----TTTAATAAATT 352  
DB 68344 ATGATTTCAAGTGAATGATTAATCTGAGTCTGATGATGATTAATTAATTTCTGTAAGT 68403  
QY 353 TTAATAACCTAATACCTTAATTTATTAAGAAGCTTTGAATGAGGCGGACGAGTAGCTC 412

```

Db      68404 TGAATATTTTCAAAATTAATTTATTTTAAAAAACCACAGGCGGCGATGTGGCTC 68463
QY      413 CTGCTGTTATCCCAACTTTGGAGGCGGCGATCCCTGAGTCCAGAGT 472
Db      68464 AGCGCTGTATCTCAACATTTTGGAGGCGCAAGGTAGTGATCACTTAGGTGAGGAT 68523
QY      473 TTAAGACGAGCTGGCGCAACTGTGAAACCTGTCTCTACTTAAAAACGCAAAATTAGC 532
Db      68524 TCGAGACAAAGCTGGCGCAACTGAGAAACCTGTCTCTACTTAAAAATTCAAAATTAGC 68583
QY      533 CAGGTGTGTGGCGATGACCTGTAGTCCCACTACTAGAGGTTGAGGAGAGATTTG 592
Db      68584 CAGGTGTGTGGCGACCGACCTGTATCTGAGCTATTGGAAGTCTGAGGCGAGAGATTTG 68643
QY      593 CTGGAACCTAGAGAGTGTGAGGTTGCGTAAACCGGATGTC-C---TGACCTCGAGCT- 648
Db      68644 CTTGAACCCAGAGGAGGAGGTTTACAGTACCGGATGCGACATTGCACTCCAGCCTG 68703
QY      649 GCGAAGAGAGCAAGACTCCATAAAGACAAACAAAGCTTGA 691
Db      68704 GGTGACAGAGCAAGACTCTGTCTCAAAACAAACAAACAAACAA 68746

RESULT 29
AC004803/c 154369 bp DNA linear PRI 25-FEB-2003
LOCUS      Homo sapiens 12 BAC RP11-355B12 (Roswell Park Cancer Institute
DEFINITION Human BAC library) complete sequence.
ACCESSION AC004803
VERSION    AC004803.11 GI:28557808
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homidae; Homo.
            1 (bases 1 to 154369)
REFERENCE  1
AUTHORS   Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
            Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Ayale, M., Banks, T.,
            Barbieri, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
            Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
            Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
            Cartron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
            Chen, G., Chen, R., Chen, Z., Chiu, D., Choudhry, I., Christopoulos, C.,
            Cleveland, C.D., Cox, C., Coyle, M.D., Dahorne, S.R., David, R.,
            Davila, M.L., Davis, C., Davy, Carroll, L., Dederich, D.A.,
            Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
            Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
            Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S.,
            Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,
            Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N.,
            Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
            Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlik, P.,
            Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M.,
            Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Hulyk, S.,
            Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y.,
            Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U.,
            King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N.,
            Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O.,
            Lien, C., Liu, J., Liu, W., Louised, H., Lozada, R.J., Lu, X.,
            Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P.,
            Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,
            McWhiney, E., McLeod, M.P., Meador, M., Mel, G., Merscher, S.,
            Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T.,
            Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, N.,
            Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
            Nguyen, N., Nickerson, E., Nwokwenko, S., Ogun, M., Okunolu, G.,
            Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
            Peters, L., Pickens, R., Primus, E., Pu, L.U., Quiles, M., Ren, Y.,
            Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Rultz, S., Saverly, G.,
            Scherer, S., Scott, G., Shen, H., Shim, C., Shoostari, N., Sisson, I.,
            Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Store, H.,
            Sutton, A., Svatek, A., Tabor, P., Tamerias, A., Tamerias, K., Tang, H.,

```

```

Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, J., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.P., Zhou, J., Zorrilla, S., Kuchelapatti, R.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 154369)
Worley, K.C.
Submitted (06-JUN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 154369)
Worley, K.C.
Direct Submission
Submitted (15-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 154369)
Worley, K.C.
Direct Submission
Submitted (25-FEB-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 25, 2003 this sequence version replaced gi:23957598.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

COMMENT
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht>

#### FEATURES

##### source

```

Location/Qualifiers
1..154369
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-355B12"
1..2210
/notes="overlaps bases 148591..150804 of clone AC004765"
/function="clone overlap"
misc_feature

```

repeat_region	357.377	/rpt_family="(T)n"
STS	1905.2020	/standard_name="G41230"
repeat_region		/complement(3042.3318)
STS	3795.4057	/rpt_family="AluSx"
repeat_region		/standard_name="G41780"
repeat_region	4718.4874	/rpt_family="AluSg/x"
repeat_region	10377.10663	/rpt_family="AluY"
repeat_region		/complement(10792.11092)
repeat_region		/rpt_family="AluSx"
repeat_region		/complement(11984.12277)
repeat_region	12434.12467	/rpt_family="AluSp"
repeat_region	12537.12610	/rpt_family="(T)n"
repeat_region		/rpt_family="Charlie1"
repeat_region	12613.12777	/rpt_family="AluJo"
repeat_region	12787.12879	/rpt_family="AluJo"
repeat_region		/rpt_family="Charlie1"
repeat_region		/complement(12924.13070)
repeat_region		/rpt_family="L1P4"
repeat_region		/complement(13098.13292)
repeat_region	13813.14115	/rpt_family="MER58A"
repeat_region		/rpt_family="AluJo"
repeat_region	14120.14220	/rpt_family="L1P4"
repeat_region	14382.14663	/rpt_family="AluY"
repeat_region	14664.14667	/rpt_family="(T)n"
repeat_region	14776.15068	/rpt_family="AluSx"
repeat_region	16255.16286	/rpt_family="AT rich"
repeat_region		/complement(16287.16598)
repeat_region		/rpt_family="AluSx"
STS	17389.17673	/standard_name="D12S888"
repeat_region	17727.17779	/rpt_family="(CA)n"
repeat_region	17863.17943	/rpt_family="(TTA)n"
repeat_region	19043.19345	/rpt_family="AluY"
repeat_region		/complement(19413.19584)
repeat_region		/rpt_family="AluJo/FRAM"
repeat_region		/complement(19588.19879)
repeat_region		/rpt_family="AluSx"
repeat_region		/complement(20140.20420)
repeat_region	20849.20875	/rpt_family="AluSx"
repeat_region		/rpt_family="(T)n"
repeat_region		/complement(20876.21127)
repeat_region		/rpt_family="AluSg"
repeat_region	22856.23154	/rpt_family="AluSp"
STS	23557.23689	/standard_name="STS-N22269"
repeat_region	23646.23685	/rpt_family="L2"
repeat_region		/complement(25629.25725)
STS	26672.26904	/rpt_family="L2"
repeat_region		/standard_name="G41789"
repeat_region	27588.27877	/rpt_family="AluJo"
repeat_region	27878.28175	

STS	/rpf family="MUSX"	30213 .30390	/standard_name="D12S1913"	31536 .31559	/rpf family="(CCA)n"	31816 .31981
STS	repeat_region					
STS						
Query Match	11.0%; Score 220; DB 8; Length 154369;					
Best Local Similarity	60.0%; Pred. No. 1.1e-35;					
Matches	420; Conservativeness 0; Mismatch 275; Indels 5; Gaps 3;					
QY	1	TGATCCACCGAGCCCTGGGCTCCCAAGTGTCTGGAGTTAAGGCGTGTAGCCACACGCTTG	60			
DB	78078	TGATCCACCTGGGCTCCGAGCTCCCAAGTGTCTGGAGTTAAGGCGTGTAGCCACACCA	78019			
QY	61	GTCCAAAGTCTTTATTTATTTGAAGACACAATGAGGAGCTTAAATCTGTCTTATTTGAC	120			
DB	78018	GCTCCAAAGGTATTTCTTATTTAAGGAGAAATTTATTTCTCTTCTTCTTATTTGGC	77966			
QY	121	AGACTTTGATGAGTCAATATCCAAATGCTGCCACTTACTGAACGCGCTTAAATGACTTAA	180			
DB	77959	TTTTTTAAACAATGGAATCTGAAGGTATGTTGGCCCAAGATTAAGACTTTTGTCTGATG	77900			
QY	181	TCTCTCAGTGTCTTTCTGCAATGTATGATGAGTGAATATATATGCTTTCAAGAGAA	240			
DB	77899	AATTTTATGCTTCTCTCTTTTATCAGTGGTCTTATGTAAGTGTGGAAGTGAAGAAA	77840			
QY	241	TAACTATGAGAAAGTGTGAGATGCTTTGATATGAAATTAAGATTTCAACAAATGAG	300			
DB	77839	GAAATGCAATTAATTAATTAAGAAATCTATTAATATGTCAGATATTTGTGCTTTATATCT	77780			
QY	301	TAGCTGCTATGTAAGATTTAAGATTTATTTATTAACAATTTAATTAATTTTAAAAAC	360			
DB	77779	CATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	77720			
QY	361	TAAATCACTTAATTTATTAAGCTTTGAATGAGGCGAGGCGAGTCTCTCTCTCT	420			
DB	77719	TTATTTATTTCTCACT-TTAAAGAAATATGAAATGAGGAGGCGGCTGCTCAACGCTCT	77661			
QY	421	AATCCCAACCTTTGGGAGGCGAAGGAGGCGGATCACTGAGGTATGAGATTTAAGAC	480			
DB	77660	AATCCCAACCTTTGGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	77601			
QY	481	AGCTGGGCAACATGTGTAAGACCTGTCTCTTAATAAAGCAAAATTTAGCAGGTGTG	540			
DB	77600	AGCTGGGCAACATGTGTAAGACCTGTCTCTTAATAAATTAAGCAAGGTGTG	77541			
QY	541	GTGGCATGCACTGTAGTCTCCAACTACTAGAGGTTGAGGAGGAGGAGGAGGAGGAGG	600			
DB	77540	GTGGCATGCACTGTAGTCTCCAACTACTAGAGGTTGAGGAGGAGGAGGAGGAGGAGG	77481			
QY	601	TAGGAGGTGAGGAGGTTGAGTAAACCGAGAT--GTCACTGCACTCAAGCTGTGCAACGA	657			
DB	77480	CGGAGAGCGAGGTTGAGTAAACCGAGAT--GTCACTGCACTCAAGCTGTGCAACGC	77421			
QY	658	GCAAGACTCCATTAAGACAAACAACTTTGAATTTGT	697			
DB	77420	AGCAAGTTCTGTCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	77381			
RESULT 30						
AC092035	162997 bp	DNA	linear	pRI 07-DBC-2001		
LOCUS						
DEFINITION	Homo sapiens chromosome 3 clone RP11-72H11, complete sequence.					
ACCESSION	AC092035					
VERSION	AC092035.2					
KEYWORDS	HTG.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Euteleostomi; Euteleostomi; Euteleostomi; Euteleostomi;					
	Homini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 162997)					

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
 TITLE Haugen, E.D.  
 JOURNAL Direct Submission  
 REFERENCE Unpublished  
 AUTHORS 2 (bases 1 to 162997)  
 TITLE Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and  
 JOURNAL Haugen, E.D.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (16-JUN-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 162997)  
 TITLE Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
 JOURNAL Haugen, E.D.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (07-DEC-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 On Dec 7, 2001 this sequence version replaced gi:14475968.  
 COMMENT ----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UMG  
 Web site: http://www.genome.washington.edu  
 Contact: uwgchtlgs@u.washington.edu  
 Drafting Center: BCM

----- Project Information  
 Center project name: chr-3  
 Center clone name: RP11-72H11 (bc0191)  
 ----- Summary Statistics  
 Sequencing vector: M13; L08821; 49% of reads  
 Sequencing vector: plasmid; L08752; 51% of reads  
 Chemistry: Dye-terminator RT; 31% of reads  
 Chemistry: Dye-terminator Big Dye; 69% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 162937 bases at least Q40  
 Consensus quality: 162997 bases at least Q30  
 Consensus quality: 162997 bases at least Q20  
 Insert size: 189816; 14.3% error; agarose-fp  
 Insert size: 170535; sum-of-contigs  
 Quality coverage: 8.5x in Q20 bases; agarose-fp  
 Quality coverage: 9.5x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:  
 5': Mapping in progress  
 3': RP11-782N4 (UMGC:bc0561) AC023231

----- Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

#### ----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

ECORI		HindIII		BglII	
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----
8696	8877	9015	9409	17217	16762
-----	-----	-----	-----	-----	-----
6	<800	6382	6539	2067	2096
-----	-----	-----	-----	-----	-----
2241	2268	512	<800	4800	5125
-----	-----	-----	-----	-----	-----
9160	8877	449	<800	392	<800
-----	-----	-----	-----	-----	-----
627	<800	7149	7144	5065	4984
-----	-----	-----	-----	-----	-----
12095	11762	10802	10131	2902	2898
-----	-----	-----	-----	-----	-----
4269	4232	172	<800	5106	5125
-----	-----	-----	-----	-----	-----
3622	3617	1312	1235	399	<800
-----	-----	-----	-----	-----	-----
76	<800	299	<800	5472	5427
-----	-----	-----	-----	-----	-----
8290	8877	574	<800	1257	1274
-----	-----	-----	-----	-----	-----
3310	3248	539	<800	15777	16762
-----	-----	-----	-----	-----	-----
1177	1194	5499	5445	913	876
-----	-----	-----	-----	-----	-----
471	<800	2632	2668	1851	1888
-----	-----	-----	-----	-----	-----
3341	3248	505	<800	302	<800
-----	-----	-----	-----	-----	-----
3806	3744	1570	1621	5971	5922
-----	-----	-----	-----	-----	-----
8735	8877	1662	1748	2385	2351
-----	-----	-----	-----	-----	-----
9467	9473	1925	1948	8553	8469
-----	-----	-----	-----	-----	-----
4680	4647	287	<800	129	<800
-----	-----	-----	-----	-----	-----
1777	1789	5559	5445	2327	2351
-----	-----	-----	-----	-----	-----
13241	12854	404	<800	4257	4244
-----	-----	-----	-----	-----	-----
12983	12854	974	891	1128	1139
-----	-----	-----	-----	-----	-----
2918	2928	1457	1446	588	<800
-----	-----	-----	-----	-----	-----
17847	18077	349	<800	2088	2096
-----	-----	-----	-----	-----	-----
2843	2802	741	<800	4796	4984
-----	-----	-----	-----	-----	-----
10470	10353	50	<800	3063	3069
-----	-----	-----	-----	-----	-----
1195	1194	2476	2519	281	<800
-----	-----	-----	-----	-----	-----
1940	1970	567	<800	2172	2223
-----	-----	-----	-----	-----	-----
4982	4986	543	<800	93	<800
-----	-----	-----	-----	-----	-----
8902	8877	1890	1948	8689	8469
-----	-----	-----	-----	-----	-----
8529	8877	4417	4370	1940	1982
-----	-----	-----	-----	-----	-----
		1391	1446	9568	9198
		10779	10131	755	<800
		1461	1446	3576	3554
		1033	975	1018	926

1823	1948	2555	2605
3444	3493	117	<800
3859	3917	2023	2096
19991	20261	7990	8018
3008	3067	1044	1050
8698	8729	4448	4427
378	<800	1507	1529
4338	4370	8086	8018
8943	8729	562	<800
1891	1948	1112	1274
2502	2519	15155	14844
408	<800		
1360	1446		
3017	3067		
2486	2519		
9784	10131		
1080	1027		
438	<800		
990	975		
1196	1122		
215	<800		

Query Match 11.0%; Score 220; DB 8; Length 162997;  
 Best Local Similarity 80.4%; Pred. No. 1.1e-35;  
 Matches 283; Conservative 0; Mismatches 65; Indels 4; Gaps 2;

QY	389	GAATGGGGCCAGGAGTACCTGCTGCTTAATCCCAACCTTTGGAGGCCAAGGTG	448
DB	37689	GATGTGGCCAGGACAAATGGCTCAAGCTCTTAATCCCAACCTTTGGAGGCCAAGGTG	37748
QY	449	GGCGGATCACTGAGGTGAGAGTTAAACCAAGCCTGGCCAAATGTGTAACCTCTTC	508
DB	37749	GGCAATCACTGAGGTGAGAGTTCAAGACCAAGCCTGGCCAAATGTGTAACCTCTTC	37808
QY	509	TCTACTAAAAAGCCAAAAATTAAGCAAGTGTGTGGCATGCACTGTAGTCCCAACTACT	568
DB	37809	TCTACTAAAAATATCAAAAAAATCTGCAAGGTGTGTGGCCGCACTGCAATCCTGTACT	37868
QY	569	CAGGAGTTGAGGGGAGGAATGTGTAAGCTAGAGGTGAGAGTTGACAGTAACCCAG	628
DB	37869	CAGGAGTTGAGGGGAGGAATGTGTAAGCTAGAGGTGAGAGTTGACAGTAACCCAG	37928
QY	629	---ATGCACTGCACTCCAGCTT-GGCAACAGAGCAAGATCCATTAAGACCAACAAAC	684
DB	37929	ATCATACCACTGCACTCCAGCTTGGCAACAGAGCAAGATCCATTAAGAAAAA	37988
QY	685	TTTGAATTTGTAAATGAGTTGTAATCTTCACTTTAAGAAATTCATCTT	736
DB	37989	AAAAAAAAAAAAAAAAAAGAGTGAAGATTTTCTATCATATATTAATGTT	38040

RESULT 31  
 AC067852/c  
 LOCUS 179146 bp DNA linear PRI 27-FEB-2003  
 DEFINITION Homo sapiens chromosome 17, clone RP11-400F19, complete sequence.  
 AC067852  
 AC067852.23 GI:28570403  
 HTG.  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 179146)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 17, clone RP11-400F19  
 Unpublished  
 2 (bases 1 to 179146)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,  
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
 Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kamm, L., Karatas, A.,  
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,  
 McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheters, R.,  
 Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, T., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Plesni, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,  
 Testafave, S., Theodore, J., Titrill, A., Travers, M., Trigglio, J.,  
 Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Direct Submission  
 Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 179146)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collymore, A., Cooke, A., Cooke, P., Corum, B., DeArrellano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
 MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunhng, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Testafave, S., Theodore, J., Topham, K., Travers, M.,  
 Vasilev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zemek, U., Zimmer, A. and Zody, M.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Direct Submission  
 Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 179146)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collymore, A., Cooke, A., Cooke, P., Corum, B., DeArrellano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,



Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, T., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhag, P., Pierre, N., Rachupka, A., Ramasamy, V., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zaimoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (27-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 27, 2003 this sequence version replaced gi:28173140.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/XM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L9335  
Center clone name: 400\_F\_19  
-----

location/Qualifiers  
1.179146  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-400F19"  
/clone\_id="RP11-400F19 Human Male BAC"  
138..440  
/rpt\_family="AluJo"  
repeat\_region  
962..1001  
/rpt\_family="GC\_rich"  
repeat\_region  
1058..1081  
/rpt\_family="GC\_rich"  
repeat\_region  
1327..1382  
/rpt\_family="CT-rich"  
repeat\_region  
2432..2469  
/rpt\_family="GC\_rich"  
repeat\_region  
2902..2934  
/rpt\_family="(AGGGGG)n"  
repeat\_region  
complement(3214..3365)  
/rpt\_family="MIR3"  
repeat\_region  
complement(5216..5317)  
/rpt\_family="MIR"  
repeat\_region  
5700..5721  
/rpt\_family="GC\_rich"  
repeat\_region  
complement(6114..6215)  
/rpt\_family="MER91A"  
repeat\_region  
7761..7848  
/rpt\_family="(CCCC)n"  
unsure  
7868..7873  
/note="<30 qual SNGL region"  
unsure  
7874..7931  
/note="<30 qual single clone coverage"  
repeat\_region  
7893..7930  
/rpt\_family="GC\_rich"  
repeat\_region  
complement(8646..8917)  
/rpt\_family="L2"  
repeat\_region  
8918..9216  
/rpt\_family="AluSc"  
repeat\_region  
complement(9217..9271)  
/rpt\_family="L2"  
repeat\_region  
9272..9569

repeat\_region  
complement(9570..9622)  
/rpt\_family="L2"  
repeat\_region  
12682..13009  
/rpt\_family="AluY"  
repeat\_region  
13015..13134  
/rpt\_family="FLAM\_C"  
repeat\_region  
13143..13185  
complement(13251..13398)  
/rpt\_family="(TA)n"  
repeat\_region  
complement(13736..14068)  
/rpt\_family="L2"  
repeat\_region  
complement(14071..14366)  
/rpt\_family="AluSc"  
repeat\_region  
complement(14367..14502)  
/rpt\_family="AluDb"  
repeat\_region  
complement(15104..15222)  
/rpt\_family="MER21B"  
repeat\_region  
15223..15277  
/rpt\_family="Tigger5"  
repeat\_region  
15247..15319  
/rpt\_family="Tigger5"  
repeat\_region  
complement(15409..15704)  
/rpt\_family="AluSc"  
repeat\_region  
complement(15710..16000)  
/rpt\_family="AluJo"  
repeat\_region  
complement(16010..16066)  
/rpt\_family="Tigger5"  
repeat\_region  
complement(16069..16668)  
/rpt\_family="MER21B"  
repeat\_region  
18847..18889  
/rpt\_family="G-rich"  
repeat\_region  
complement(18986..19008)  
/rpt\_family="L1MB3"  
repeat\_region  
complement(19009..19048)  
/rpt\_family="L1MB3"  
repeat\_region  
complement(19049..19352)  
/rpt\_family="AluJo"  
repeat\_region  
complement(19353..19508)  
/rpt\_family="L1MB3"  
repeat\_region  
complement(19509..19929)  
/rpt\_family="L1MB3"  
repeat\_region  
19931..20323  
/rpt\_family="L1MB3"  
repeat\_region  
complement(20327..20627)  
/rpt\_family="AluSc"  
repeat\_region  
complement(20645..20929)  
/rpt\_family="AluSp"  
repeat\_region  
complement(20930..21218)  
/rpt\_family="AluSc"  
repeat\_region  
complement(21223..21322)  
/rpt\_family="AluSc"  
repeat\_region  
complement(21323..21614)  
/rpt\_family="AluJo/F(R)AM"  
repeat\_region  
complement(21615..22365)  
/rpt\_family="AluSc"

Query Match 11.0%; Score 220; DB 8; Length 179146;  
Best Local Similarity 60.8%; Pred. No. 1,1e-35;  
Matches 435; Conservative 0; Mismatches 265; Indels 16; Gaps 4;

Db 1 TGATCCACGACCTGGCCCTCCCAAGTGTGGATTACAGCGGTGACGACGACGCTG 60  
43622 TGATCCGCTGCTCAGCCCTCCCAAGTGTGGATTACAGCGGTGACGACGACG 43563  
61 GTGGAATGCTTATTAATTTGAAGACCAATGCGCTTAATCTGCTTATTTGAC 120  
43662 GCTTACAGTGT---TGATTTAACAAACCAATGATAGTAGAGAACTAAATCATGTCAA 43506  
121 AGACTTTGATGAGTCAATCCCAATGCTGCCACTTACTGAAGGCTTAATGACTTAG 180  
43505 TGTTAATGCTTAATTAACAAATATCTTAATGTTTAAATTTGCTTAAAGCACTG 43446

QY 181 TCTCTCAGCTGCTCTTTCGATAT-----GTAAAGTGAATTAATGAGCTTTTCA 232  
 DB 43445 TGTACTAAGCTGTAGTCTTACTACTAGAGGCTGAGCAAGAAATGCTTCAGC 43386  
 QY 233 AGAGGAAATTAACCTTAAGAAAGTTGAGGATAGTCTTGAATATGAATTAAGATTTC 292  
 DB 43385 CCCAGAGATTCAAGGTTGAGTGAAGTATATGATGACATCCACTCCAGTCTGGGCAAC 43326  
 QY 293 ACAAGTAGTAGCTGCTATGTAAGATTTAAGATTTATTTACATCAATTAATTAATTAAT 352  
 DB 43325 AGAGAGCTCTGTTTCTTACAAAATAATTAATTAATTAATTAATTAATTAATTAATTA 43266  
 QY 353 TTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 412  
 DB 43265 TGAGCTAAATTTCTCAACATATCAATTAATTAATTAATTAATTAATTAATTAATTA 43206  
 QY 413 CTGCTCTTAATCCCAACATTTTGGAGGCCAAGGTGGCGGATCAGTGAAGTCAAGAGT 472  
 DB 43205 ACCGCTGTATCCAGACATTTTGGAGGCCAAGATGGCGGATCAGTGAAGTCAAGAGT 43146  
 QY 473 TTAAGACCAAGCTGGCCAAATGATGTAAACCTGTCTACTAATAAAGCAAAATTAAGC 532  
 DB 43145 TCAAGACCAAGCTGGCCAAATGATGTAAACCTGTCTACTAATAAAGCAAAATTAATC 43086  
 QY 533 CAGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 592  
 DB 43085 TGGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 43026  
 QY 593 CTTGAACCTTAAGAGTGTGAGGATTTGACATTAACCCGAGATGTAC--TGACATCCAGCTG 649  
 DB 43025 CTTGAACCTTAAGAGTGTGAGGATTTGAGTGTGATGATGATGATGATGATGATGATG 42966  
 QY 650 G--CAACAGAGCAAGCTCTATTAAGCAACAAAGTTGAATTTGTGAATGA 703  
 DB 42965 GCGACAGAGCAAGCTCTATCTCAAAAATTAATTAATTAATTAATTAATGA 42910

RESULT 32  
 BV651164/c 841 bp DNA linear STS 16-Ap-2005  
 LOCUS S21P6161RA5.T0 Noemie Pan troglodytes troglodytes STS genomic,  
 DEFINITION sequence tagged site.  
 ACCESSION BV651164  
 VERSION BV651164.1 GI:62679133  
 KEYWORDS STS.  
 SOURCE Pan troglodytes troglodytes  
 ORGANISM Pan troglodytes troglodytes  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Pan. 1 (to 841)  
 REFERENCE 1 (bases 1 to 841)  
 AUTHORS Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and  
 Jaffe,D.B.  
 TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the  
 Human Genome  
 JOURNAL Unpublished (2005)  
 COMMENT  
 Contact: Michael C. Zody  
 Broad Institute of MIT and Harvard  
 320 Charles Street, Cambridge, MA 02141, USA  
 Tel: 6172580933  
 Fax: 6172580903  
 Email: mczody@broad.mit.edu  
 Primer A: No sequence submitted  
 Primer B: No sequence submitted  
 STS size: 841  
 Protocol:  
 23,021,928 chimpanzee whole genome shotgun reads were aligned to  
 the Human genome NCBI  
 Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,  
 including Clint (Pan  
 troglodytes verus), 3 other Pan troglodytes verus chimps  
 (Donald, Karlén, Yvonne), 3 Pan

troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps  
 of unknown origin  
 (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the  
 western chimp and Pan  
 troglodytes troglodytes is the central chimp. To be included in  
 chimpanzee SNP discovery, a  
 read must be at least 500bp in length, at least 50% of its base  
 calls must have Phred  
 score >= 20, at least 30% of its base calls must satisfy  
 SNOS(30,25) (single strand NOS, the  
 base in question has Phred score >= 30, the surrounding 10 bases in  
 the read have Phred  
 score >= 25), and the read must have at least 200 bp SNOS(30,25)  
 bases. Reads not uniquely  
 placed in the genome and read pairs whose two ends were not  
 consistently placed were  
 discarded. After above filtering, NOS(30,25) standard was applied  
 to all pairs of  
 overlapping reads to call NOS bases and SNPs. Alignments (between  
 two reads) with less  
 than 100 NOS bases or with SNP rate > 0.01 were discarded. To  
 exclude alignment between two  
 copies of a single read, comparisons between two reads that share  
 95% of their genome  
 alignments (>=95% bases of read A and >=95% bases of read B were  
 placed at the same locus  
 of human genome) were discarded.

## FEATURES

source  
 1. 841  
 /organism="Pan troglodytes troglodytes"  
 /mol\_type="genomic DNA"  
 /sub\_species="troglodytes"  
 /db\_xref="taxon:37011"  
 /clone\_lib="Noemie"  
 <1..>841

## ORIGIN

Query Match 11.0%; Score 219.2; DB 10; Length 841;  
 Best Local Similarity 70.1%; Pred. No. 4,4e-35;  
 Matches 324; Conservative 0; Mismatches 133; Indels 5; Gaps 2;  
 QY 215 GAATATGATGCTTCAAGAGGATTAACCTATGAAAGTGTGAGATAGTGTTCGA 274  
 DB 464 GAGCAAGACTCGGTCTCAAAAAAAGAAATAATCAAAAATGTTAGAAATAAGA 405  
 QY 275 TATGAATAAGATTCAACAGTAGTACTGCTATGAAGTTAAGGT--TATTTAT 332  
 DB 404 AGGACATTATATATTAATTAAGTTCAATATTAAGAAAGATATTAACAATTGTAAATAT 345  
 QY 333 TACAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 392  
 DB 344 TTAACACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 285  
 QY 393 TGGGCGAGGCGAGTAGTCTGCTGTATCTCCCAACATTTTGGAGGCCAAGTGGCG 452  
 DB 284 AAGCCAGGCGAGTGTCTATGATCTGATATCTCAGACATTTTGGAGGCCAAGTGGCG 225  
 QY 453 GATCAGCTGAGGTCAAGAGTTTAAGCAAGCTGTGGCAACATGTGAACCTGTCTTA 512  
 DB 224 GATCAGCTGAGGTCAAGAGTTTAAGCAAGCTGTGGCAACATGTGTCTTA 165  
 QY 513 CTTAAAAACCAAAAAATTAAGCAAGTGTGTGGCATGACCTGTAGTCCCACTACTCAG 572  
 DB 164 CTTAAAAATCAAAAAATTAAGCAAGTGTGTGGCATGACCTGTAGTCCCACTACTCAG 105  
 QY 573 AGCTGAGGAGAGAGATTGCTTGAACCTTAGAGGTGAGGTTGCAAGTGAACCCGAGA--- 629  
 DB 104 AGACTGAGTCAGAGAGATGCTGAACCCAGAGGTGAAGTTGCAAGTGAACAGAGATG 45  
 QY 630 TGTCACTGACTTCAGCTGGCAACAGAGAGATCTCAATA 671  
 DB 44 TGCACATGACATCCAGCTGGCAACAGAGAGATCTCAATA 3

**RESULT 33**

LOCUS	ALJ59713	152966 bp	DNA	linear	PRI 18-MAY-2005
<b>DEFINITION</b>	Human DNA sequence from clone RP11-95P3 on chromosome 6 Contains the 5' end of the RU2 gene, the MRS2L gene for MRS2L-like magnesium homeostasis factor; the 3' end of the GPLD1 gene for glycosylphosphatidylinositol specific phospholipase D1 and two Cpg islands, complete sequence.				
<b>ACCESSION</b>	ALJ59713				
<b>VERSION</b>	ALJ59713 .25 GI:13938809				
<b>KEYWORDS</b>	HTG; Cpg island; glycosylphosphatidylinositol specific phospholipases; GPLPD; GPLD1; HPT; magnesium homeostasis factor; MRS2L; phosphatidylinositol-glycan-specific phospholipase; PIGPLD; PIPLD1; RU2; RIZAS; RU2S.				
<b>SOURCE</b>	Homo sapiens (human)				
<b>ORGANISM</b>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.				
<b>REFERENCE AUTHORS</b>	Kimberley, A. Direct Submission Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On May 3, 2001 this sequence version replaced gi:13446455.				
<b>JOURNAL</b>	The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMEP; Information on the WORMEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormep">http://www.sanger.ac.uk/Projects/C_elegans/wormep</a> This sequence was generated from part of bacterial clone conlifs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <a href="http://www.sanger.ac.uk/HGP/Chr6">http://www.sanger.ac.uk/HGP/Chr6</a> RP11-95P3 is from the library RPc1-11.1 constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a> VECTOR: pBACE3.6				
<b>COMMENT</b>	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: <a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a> Contact: vegas@sanger.ac.uk ----- This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. Location/Qualifiers 1..152966 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="6" /clone="RP11-95P3" /clone_1ib="RPc1-11.1" 1 /note="Clone_left_end: RP11-95P3" 17700 /note="Clone_right_end: RP11-40B20" complement(49750) complement(join(49751..49838,50088..50142,53977..>54172)) /gene="DCCDC2" /locus_tag="RP11-26OH5_B.1-003" complement(join(49751..49838,50088..50142,53977..>54172)) /gene="DCCDC2" /locus_tag="RP11-26OH5_B.1-003" /product="doublecortin domain containing 2"				

  

gene	polyA_signal	CDS
gene	polyA_site	CDS
<pre> complement(49765..49770) /gene="DCCDC2" /locus_tag="RP11-26OH5_B.1-003" complement(join(49824..49838,50088..50142,53977..&gt;54172))) /gene="DCCDC2" /locus_tag="RP11-26OH5_B.1-003" standard_name="OTTHUMP0000017889" /codon_start=3 /product="doublecortin domain containing 2" /protein_id="CAI17105.1" /db_xref="GI:55959678" /db_xref="GOA:Q5Y3Y5" /db_xref="InterPro:IPR003533" /db_xref="UniProt/TREMBL:O5T3Y5" /tranlation="VVIHEKKVSFEFVEIKEVGVQAPGAVRNIYTPRTGRIRRLDQIQSGNYVAGCGDEAFKXNLYLDIGEIKRPMHVTEILHYV" join(complement(53977..54571),complement(50088..50142),complement(ALJ59389.15:80898..80974),complement(ALJ59389.15:80645..80776),complement(ALJ59389.15:69862..70008),complement(ALJ59389.15:67782..67836),complement(ALJ59389.15:66979..57141),complement(ALJ591344.1:45608..45708),complement(ALJ591344.1:18936..19228),complement(ALJ591344.1:12362..15440)) /gene="DCCDC2" /locus_tag="RP11-26OH5_B.1-001" join(complement(53977..54571),complement(50088..50142),complement(ALJ59389.15:80898..80974),complement(ALJ59389.15:80645..80776),complement(ALJ59389.15:69862..70008),complement(ALJ59389.15:67782..67836),complement(ALJ59389.15:66979..57141),complement(ALJ591344.1:45608..45708),complement(ALJ591344.1:18936..19228),complement(ALJ591344.1:12362..15440)) /gene="DCCDC2" /locus_tag="RP11-26OH5_B.1-001" /product="doublecortin domain containing 2" /note="match: ESTs: AA004587,AA004717,AT889161,AL564319,AV750645,BF593580,BF876266,BT760196,BM468677 match: CDNs: AF181721.1 AK027036.1 BC014954.1" join(complement(53977..54269),complement(50088..50142),complement(ALJ59389.15:80898..80974),complement(ALJ59389.15:80645..80776),complement(ALJ59389.15:69862..70008),complement(ALJ59389.15:67782..67836),complement(ALJ591344.1:45608..45708),complement(ALJ591344.1:18936..19228),complement(ALJ591344.1:15336..15440)) /gene="DCCDC2" /locus_tag="RP11-26OH5_B.1-001" /standard_name="OTTHUMF0000017887" /note="match: proteins: Q9N080,Q9UHG1,Q9ULR6" /codon_start=1 /product="doublecortin domain containing 2" /db_xref="GI:55959677" /db_xref="GOA:Q9UHGO" /db_xref="UniProt/Swiss-Prot:Q9UHGO" /tranlation="MSGGSARSHLSQPVAKSVALTYRNNGDPFYAGRNVIIHEKKVSSF EYFLKEVTGVQAPGAVRNIYTPRTGRIRKLDDIQSGNYVAGCGDEAFKXNLYLDI GEIKRPMHVTEVKPVHSRINVSAFRFKLPDCTFILANGDLINPASRLLIPIR KTIINOWHYLCWYTEKITLRISGAHRITLTKEGLVESGELENOGFYAVADRFCKR PYSELFDKSTKRPRPGQAKSLPIIVTSRKSKSGGNRHKSTSYGSDNSPODKR KKKEDEVSEKTTLKQNVKLNKSOETIPNSDBEGIFKAQEKSRTRGAAAYOEEDDTQ VAVPVDOFAELIVDEIDEKANKDAEKDFESGNNGULEEGEGEAIDADBPQVEIL DHSSEOQAPPARVNGGTDEENGEELQQVVNNILQLVLDERKSKSQAGSGADEADVDPOR PRPEYKITSPEENENNQQOKDYAAVA" 53422..54803 /gene="KNA61" </pre>		

	mRNA	/locus_tag="RP11-95P3.2-001"	53422..54803	/gene="KXAG1"	/locus_tag="RP11-95P3.2-001"	/product="kidney associated antigen 1"	/note="match: cDNAs: AF181722.1"	54159..54413	/gene="KXAG1"	/locus_tag="RP11-95P3.2-001"	/standard_name="ORTHUMP00000016084"	/note="match: proteins: Q9UBB8"	/codon_start=1	/product="kidney associated antigen 1"	/protein_id="CAI17106.1"	/db_xref="GI:55959679"	/db_xref="Genew:21031"	/db_xref="GOA:Q9UBB8"	/db_xref="UniProt/TREMBL:O9UBP8"	/translation="MDDDAAPRVGEPVAVHKAHLHDGLRQVNGPGAAAAHLPRWPP QLAASRRARPPLSQRHRTQGAGSPPEINTEKLTPVKER"	54782..54787	/gene="KXAG1"	/locus_tag="RP11-95P3.2-001"	54804	polyA_site	misc_feature	digest."	/note="Tandem repeat. Sequence from overlapping clone dj282H10 (AL132672). Assembly confirmed by restriction digest."	join(199473..99755,101687..101760,104927..104963, 105980..106092,108741..108914,111552..111682, 112916..113032,114603..114755,114980..115097, 119456..119569,120103..122101) /locus_tag="RP11-95P3.1-009"	join(199473..99755,101687..101760,104927..104963, 105980..106092,108741..108914,111552..111682, 112916..113032,114603..114755,114980..115097, 119456..119569,120103..122101) /locus_tag="RP11-95P3.1-009"	/note="match: ESTs: AA280673 AA781573 AA782213 AA890448 AA931950 AA894525 AI027389 AI342569 AI522078 AI525690 AI628970 AI954726 AV712957 AM131633 AM295997 AM632241 AM974621 BE168047 BE219330 BE463602 BE798677 BF116203 BF216598 BF218387 BF245307 BF248875 BG421926 BG740291 BM018867 match: cDNAs: AF288288.1 AF288289.1 AF288290.1 AF288291.1 AF293076.1"	join(199473..99755,101687..101760,104927..104963, 105980..106092,108741..108914,111552..111682, 112916..113032,114603..114755,114980..115097, 119456..119595) /locus_tag="RP11-95P3.1-011"	join(199473..99755,101687..101760,104927..104963, 105980..106092,108741..108914,111552..111682, 112916..113032,114603..114755,114980..115097, 119456..119595) /locus_tag="RP11-95P3.1-011"	/note="match: ESTs: A1188501 A1206137 A1307274 A1624310	Query Match 11.0%; Score 219.2; DB 8; Length 152966; Best Local Similarity 59.8%; Pred. No.1,6e-35; Matches 427; Conservative 0; Mismatches 273; Indels 14; Gaps 3;	OY 1 TGATCCACGAGCTTGCGCTCCCAAGTCGTGGAGTTACAGGCGTAGGCCACGCGCTG 60 DB 130251 TGAATCACCCACTCAGCTCCGACCAAGTCTGGAGATTACGGGATTAAGCCACGACCACA 130310 OY 61 GTCGAATGCTTTATTATTGGAAGACACAACATGGGCCTTAAATCTGCTCTTATTATGAC 120 DB 130311 GGCTCATTTCTACATTTCTTGATTAGCTTACAAATGTGCATCCCTTTGACCCCCAGAAA 130370 OY 121 AGACTTTGATGAGATCCAATCCCAATGCTGCGCCTTACGATCGAACGGCGCTTAAATGACTTGA 180 DB 130371 TTCTACTTATATAAATATACGGGTGAGAAAATAACAGATGTATATAAAACATTTATGTGAC 130430
--	------	------------------------------	--------------	---------------	------------------------------	--	----------------------------------	--------------	---------------	------------------------------	-------------------------------------	---------------------------------	----------------	--	--------------------------	------------------------	------------------------	-----------------------	----------------------------------	---	--------------	---------------	------------------------------	-------	------------	--------------	----------	---	---	---	---	--	--	---	---	--

QY	181	TCCTCTCAGAGCTGCTCTTCTGCAATGTGAAGTGAATATATATGAGCTTTCAAGAGGAA	240
Db	130431	TATTTTACCAATATATATTTAAACAATGAANAATTGGAAATGSCACATTTAGCTATCAATGGT	1304300
QY	241	TAAACCTATGAAAGTGTGAGAGATAGCTTTGATATGANAATTAAGATTTCAACAAGTAG	300
Db	130491	TGCCGAAATTAATGGGTAAATAAACAATATGGTTGTGTGATATAAATACTCATACAGCCGG	1305560
QY	301	TAGCTGCTATGAGATTTAGAGTTATTTATTACACATATTTTAATTAATTTTAAAAAC	360
Db	130551	TAAATATCATATTTATTAATGAATATCAGATTAACATPAGAAAATGTTCTCGGT	1306020
QY	361	TAATACCTTAATTAATTTAAAGAGCTTTGAATATGGGCGCCAGGCGAGTAGCTCTGCTGT	420
Db	130603	--ATATTTGTATAGTGAAGAAAAGAAAAATCAAGTTACGGCCAGGCAACGTGCTCACACTAT	1306660
QY	421	AATCCCAACACTTTTGGGAGGCCAAGTGGCGGATCACCCTGAGTCAGAGATTTTAAGACC	480
Db	130661	AATCCACAGACTTTTGGAGGCTGAGAGCGGAGATCACTGAGAGTCAAGATTTGAGACC	1307200
QY	481	AGCTTGGCCAAACATGTGTGAACCTCTGTCTTACTTAATAAAGCAAAATTAAGCAGGTG	540
Db	130721	AGCTTGGCCAAACGTGTGAACCCCGTCTTACTTAATAAAGCAAAATTAAGCAGGACAG	1307800
QY	541	GTGGCATGCACCTGTATGTGCCAATCACTACAGAGGTGAGGGAGGAGATTTGCTGAACC	600
Db	130781	GTGGGTGTGCTGTATGTATGCCACTTACAGAGAGTGAAGAGGAGGAGATTCGTTGAACC	1308400
QY	601	TAGAGAGTGAAGTGTGACAGTAACCCGAG--ATGTCACTGCACTCCAGCCTG3-CAMCAG	656
Db	130841	TGGAGAGTGAAGGTTTACTAGAGCCGAGATCATGCACTGCAGCTCCAGCTGAGATGACAG	1309000
QY	657	AGCAAGACTCCATTAAGACACAAAGCTTTGAATTTGTATTAATGAGTTGAC	710
Db	130901	AGTGAAGCTCCATTAAGAAAAAAGTTACAAAAATTAAGCAAAATTAATGAC	1309540
RESULT 34			
AC027113			
LOCUS	AC027113	159391 bp	DNA linear HTG 24-AUG-2002
DEFINITION	Homio sapiens chromosome 1 clone RP11-67212 map 1, WORKING DRAFT		
ACCESSION	AC027113.3	GI:8081286	
VERSION	AC027113.3		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homio sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 159391)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,		
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
JOURNAL	Homio sapiens chromosome 1, clone RP11-67212		
REFERENCE	2 (bases 1 to 159391)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,		
	Anderson,S., Baldwin,J., Berna,N., Bastien,V., Bede,F.,		
	Boguslavsky,L., Bourhgalter,B., Brown,A., Burkett,G.,		
	Campopiano,A., Castle,A., Chopelet,Y., Colangelo,M., Collins,S.,		
	Collamore,A., Cooke,P., Dextrallano,K., Dewar,K., Diaz,J.S.,		
	Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,		
	Galagan,J., Gardyna,S., Grinde,S., Goyette,M., Graham,L.,		
	Grand-Pierre,N., Grant,G., Hagos,B., Healdorf,A., Horton,L.,		
	Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,		
	Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,		
	Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,		
	McCarthy,M., McKwan,P., McGurk,K., McKernan,K., McPheters,R.,		
	Meldrum,J., Menes,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,		
	Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,		
	O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,		
	Plauti,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,		
	Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,		

**TITLE** Direct Submission  
**JOURNAL** Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**REFERENCE** 3 (bases 1 to 159391)  
**AUTHORS** Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

[illegible]

RESULT 35  
AC017038  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
TITLE  
AUTHORS  
JOURNAL  
COMMENT

AC017038 180707 bp DNA linear HTG-07-JUL-2000  
Homo sapiens chromosome 8 clone RP11-311B2, WORKING DRAFT SEQUENCE.  
20 unordered pieces.  
AC017038  
AC017038.6 GI:8568131  
HTG: HTGS PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
Homo sapiens  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.  
1 (bases 1 to 180707)  
Waterston, R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 180707)  
Waterston, R.H.  
Direct Submission  
Submitted (09-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jun 16, 2000 this sequence version replaced gi:7230978.

```

Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H NH031B02
Summary Statistics -----
Sequencing vector: M13; %
Sequencing vector: plasmid; %
Chemistry: Dye-primer ET; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 168786 bases at least Q40
Consensus quality: 172458 bases at least Q40
Consensus quality: 174496 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 178807; sum-of-contrigs
Quality coverage: 3.48 in Q20 bases; sum-of-contrigs
Quality coverage: 3.48 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2076: contig of 2076 bp in length
2077
2176: gap of unknown length
2177
4769: contig of 2593 bp in length
4770
4669: gap of unknown length
4870
7558: contig of 2689 bp in length
7559
7658: gap of unknown length
7659
10945: contig of 3287 bp in length
10946
11045: gap of unknown length
11046
13945: contig of 2900 bp in length
13946
14045: gap of unknown length
14046
17724: contig of 3679 bp in length
17725
17824: gap of unknown length
17825
22156: contig of 4332 bp in length
22157
22256: gap of unknown length
22257
26042: contig of 3786 bp in length
26043
26142: gap of unknown length
26143
32662: contig of 6520 bp in length
32663
32762: gap of unknown length

```

```
* 32763 41257: contig of 8495 bp in length
* 41258 41357: gap of unknown length
* 41358 47568: contig of 6211 bp in length
* 47569 54391: gap of unknown length
* 54392 54491: gap of unknown length
* 54492 62210: contig of 7719 bp in length
* 62211 62310: gap of unknown length
* 62311 70009: contig of 7659 bp in length
* 70010 70109: gap of unknown length
* 70110 78968: contig of 8859 bp in length
* 78969 90881: contig of 11813 bp in length
* 90882 90981: gap of unknown length
* 90982 105415: contig of 14434 bp in length
* 105416 105515: gap of unknown length
* 105516 121842: contig of 16327 bp in length
* 121843 141936: gap of unknown length
* 141937 142036: contig of 19994 bp in length
* 142037 180707: contig of 38671 bp in length.
FEATURES
source
1..180707
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-313B2"
1..2076
/note="assembly_name:Contig8"
2077..2176
/estimated_length=unknown
2177..4769
/note="assembly_name:Contig9"
4770..4869
/estimated_length=unknown
4870..7558
/note="assembly_name:Contig10"
7559..7658
/estimated_length=unknown
7659..10945
/note="assembly_name:Contig11"
10946..11045
/estimated_length=unknown
11046..13945
/note="assembly_name:Contig12"
13946..14045
/estimated_length=unknown
14046..17724
/note="assembly_name:Contig13"
17725..17824
/estimated_length=unknown
17825..22156
/note="assembly_name:Contig14"
22157..22256
/estimated_length=unknown
22257..26042
/note="assembly_name:Contig15"
26043..26142
/estimated_length=unknown
26143..32662
/note="assembly_name:Contig16"
32663..32762
/estimated_length=unknown
32763..41257
/note="assembly_name:Contig17"
41258..41357
/estimated_length=unknown
41358..47568
/note="assembly_name:Contig18"
47569..47668
/estimated_length=unknown
47669..54391
misc_feature
```

```
/note="assembly_name:Contig19"
54392..54491
/estimated_length=unknown
54492..62210
/note="assembly_name:Contig20"
62211..62310
/estimated_length=unknown
62311..70009
/note="assembly_name:Contig21"
70010..70109
/estimated_length=unknown
70110..78968
/note="assembly_name:Contig22"
78969..79068
/estimated_length=unknown
79069..90881
/note="assembly_name:Contig23"
90882..90981
/estimated_length=unknown
90982..105415
/note="assembly_name:Contig24"
clone_end:SP6
vector_side:right"
105416..105515
/estimated_length=unknown
105516..121842
/note="assembly_name:Contig25"
121843..121942
/estimated_length=unknown
121943..141936
/note="assembly_name:Contig26"
141937..142036
/estimated_length=unknown
142037..180707
/note="assembly_name:Contig27"
misc_feature
```

## ORIGIN

```
Query Match 11.0%; Score 219.2; DB 14; Length 180707;
Best Local Similarity 72.1%; Pred. No. 1.6e-35;
Matches 300; Conservative 0; Mismatches 113; Indels 3; Gaps 1;
```

```
QY 351 TTTTAAAACTAATACCTTAATTTTAAAGAGCTTGAAGTGGCGCCGCGCTAGC 410
Db 139975 TCTGTAAATTAAACTATCTAAGATTAAATTTCTTAAAGGCGCGCACTAGC 140034
QY 411 TCCTGCTGTAAATCCCAACTTTGGAGGCCAAGGTGGCGGATCACTGAGTCAGGA 470
Db 140035 TCATGCTGTAAATCCCAACTTTGAGAGGCTGAGGTGGATCAACGGAGTCAGGA 140094
QY 471 GTTTAAGACGAGCTGGCCAACTGTGTAACCCCTGCTCTAATAAAACGCAAAATTAA 530
Db 140095 GTTCAAGACCACTCTGGCCAACTGTGTAACCCCTGCTCTAATAAAATTA 140154
QY 531 GCCAGGTGTGTGTCATGACCTGTATGCCCACTACTAGAGAGTTGAGGAGGAGAA 590
Db 140155 GCTGGGCAATGTGGCAGTGTCTATATCCCACTCTCGGGAAGCTGAGGCGAGAA 140214
QY 591 TGCTTGAACCTTAGAGTGGAGTTGCAAGAACCGAGAA--TGTCACTGCACCTCCAGCC 647
Db 140215 CACTTGAACCCAGAGGTGAGGTGCAAGAACCGAGAACTGCACTGCACCTCCAGCC 140274
QY 648 TGGCAACAGACGAAGCTCATTAAGACCAAAAGCTTTGAATTGTGTAATGAGTTG 707
Db 140275 TGGCGACAGCAAGACTCATTCACAAAATAAAATAAAATAAAATAAAGTTTA 140334
QY 708 TACCTAATCTTCAATTAAGAAATTCATCTTGTCTCAATTAATTTTACTTGACATGAG 763
Db 140335 TTTAAGAAATAAGTTAATATCTATCCCTCAATAAATTCATTAATCTTCGTGAG 140390
RESULT 36
AC025682/c AC025682 158135 bp DNA linear PRI 01-FEB-2002
LOCUS
```



	DEFINITION	Homio sapiens chromosome 17, clone RP11-560116, complete sequence.
	ACCESSION	AC025682
	VERSION	AC025682.6 GI:18377318
	KEYWORDS	Htg.
	SOURCE	Homio sapiens (human)
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	TITLE	1 (bases 1 to 158135)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.	
JOURNAL	Homio sapiens chromosome 17, clone RP11-560116	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 158135)	
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Bana,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Bugalski,T., Meheus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pianti,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stoianovic,N., Subramanian,A., Talmas,J., Teefaye,S., Theodore,J., Tittrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
REFERENCE	3 (bases 1 to 158135)	
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Bana,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A., Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Girde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kelle,C., Larocque,K., Lamazares,R., Landers,T., Lehoccky,J., Landers,T., Lehoccky,J., Levine,R., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McSheeters,R., Meldrum,J., Meheus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pianti,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stoianovic,N., Subramanian,A., Talmas,J., Teefaye,S., Theodore,J., Tittrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
REFERENCE	On Jan 27, 2002 this sequence version replaced gi:18370031.	
COMMENT	All repeats were identified using RepeatMasker: Smt, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a>	
TITLE	Genome Center	
JOURNAL	Center: Whitehead Institute/ MIT Center for Genome Research	
COMMENT	Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a> Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a>	

```

----- Project Information
Center project name: L8312
Center clone name: 580_1_16
-----
FEATURES
Source
-----
Location/Qualifiers
1. .158135
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-580116"
/clone_1kb="RP11 Human Male BAC"
repeat_region complement(4. .215)
/rpt_family="Aluub"
repeat_region complement(237. .484)
/rpt_family="Aluub"
repeat_region complement(500. .801)
/rpt_family="AluSg"
repeat_region complement(2336. .2640)
/rpt_family="AluSx"
repeat_region .3231
/rpt_family="MERSA"
/rpt_family="MERSA"
repeat_region 3232. .3501
/rpt_family="AluSg1"
repeat_region 3502. .3588
/rpt_family="MERSA"
repeat_region 5709. .5744
/rpt_family="(A)n"
repeat_region 6191. .6237
/rpt_family="G-rich"
unure 6197. .6204
unure /note="<30 qual SINGL region"
6206. .6211
/note="probably GGGGGG, possibly GGGGG"
repeat_region 6537. .6552
/rpt_family="G-rich"
unure complement(6636. .6640)
/note="<30 qual SINGL region"
complement(6676. .6681)
/note="<30 qual SINGL region"
6847. .6873
/rpt_family="(GGGA)n"
repeat_region complement(7516. .7665)
/rpt_family="MIR3"
repeat_region complement(7646. .7737)
/rpt_family="MIR"
repeat_region 8051. .8345
/rpt_family="AluSc"
repeat_region 8360. .8463
/rpt_family="FLAM_C"
repeat_region 8504. .8807
/rpt_family="AluSx"
repeat_region 9207. .9218
/rpt_family="AluS"
repeat_region 9219. .9511
/rpt_family="AluSx"
repeat_region 9512. .9574
/rpt_family="AluS"
repeat_region 9575. .9627
/rpt_family="Alu"
repeat_region 9640. .9917
/rpt_family="AluSx"
repeat_region complement(10167. .10285)
/rpt_family="AluSg/x"
repeat_region complement(10396. .10571)
/rpt_family="AluSg"
repeat_region 10611. .10708
/rpt_family="FRAM"
repeat_region 10865. .11064
/rpt_family="MR"
repeat_region 11052. .11348
/rpt_family="AluSx"

```

```

repeat_region      11358..11387
                    /rpt_family="AT_rich"
repeat_region      complement(11592..11957)
                    /rpt_family="L2"
repeat_region      complement(11999..12068)
                    /rpt_family="Alu"
repeat_region      complement(12276..12520)
                    /rpt_family="Aluub"
repeat_region      complement(12734..13039)
                    /rpt_family="Aluub"
repeat_region      complement(13520..13787)
                    /rpt_family="AluSx"
repeat_region      complement(13792..13986)
                    /rpt_family="MER20"
repeat_region      complement(14103..14403)
                    /rpt_family="AluSx"
repeat_region      complement(14405..14714)
                    /rpt_family="Aluub"
repeat_region      complement(14715..14847)
                    /rpt_family="MIR"
repeat_region      14848..15167
                    /rpt_family="AluSg"
repeat_region      complement(15168..15300)
                    /rpt_family="MIR"
repeat_region      15415..15724
                    /rpt_family="AluSx"
repeat_region      complement(15756..15884)
                    /rpt_family="L1ME4A"
repeat_region      complement(15897..16025)
                    /rpt_family="L1ME2"
repeat_region      complement(16051..16245)
                    /rpt_family="MER63A"
repeat_region      complement(16251..16675)
                    /rpt_family="L1ME1"
repeat_region      complement(16676..16938)
                    /rpt_family="Aluub"
repeat_region      complement(16945..17019)
                    /rpt_family="L1ME1"
repeat_region      complement(17073..17206)
                    /rpt_family="L1ME1"
repeat_region      complement(17207..17511)
                    /rpt_family="Aluub"
repeat_region      complement(17512..17783)
                    /rpt_family="L1ME1"
repeat_region      complement(17784..18053)
                    /rpt_family="L1ME1"
repeat_region      complement(18054..18318)
                    /rpt_family="AluSx"
repeat_region      complement(18319..18615)
                    /rpt_family="L1ME1"
repeat_region      complement(18616..18698)
                    /rpt_family="Alu"
repeat_region      18717..18888
                    /rpt_family="L1ME1"

```

Query Match 10.9%; Score 219; DB 8; Length 158135;  
 Best Local Similarity 70.5%; Pred. No. 1.8e-35;  
 Matches 321; Conservative 0; Mismatches 130; Indels 4; Gaps 2;

```

QY 272 TGAATGAAATTAAGAGTTTCAACAAGTAGTGTCTATTGAAGATTAAAGATTATTTA 331
DB 14513 TGGTTAAAGAGTTTCAACAAGACTATTTGAACCTAGACATTAGTCTCTAGAGCTTACA 145554
QY 332 TTACAACTATTATTAATAAATTTTAAAACTAATACACTTAATATTATTAAGAGCTTTGAA 391
DB 14553 CTCACATCCCTTACAGACACTCTCAGCTAGTCTACTTAGTTCATATCAAAATTTAAAT 145494
QY 392 AT-GGGCGAGCGGAGTAGTCTCTGCTGTAACTCCCAACTTTGGAGCGCAAGTGGG 450
DB 145493 CTAAAGCGGAGCGGAGTAGTCTCTGCTGTAACTCCCAACTTTGGAGCGGCGGG 145434
QY 451 CGGATCAGCTGAGGTCAGAGTTTAAGACGAGCTGGCCAACTGTTGAACCTGTTCTC 510
DB 145433 AGGATCAGCTGAGGTCAGAGTTTAAGACGAGCTGGCCAACTGTTGAACCTGTTCTC 145374

```

```

QY 511 TACTAAAAACGCAAAATTAAGCCAGGTGTGTGCGACTGACCTTGTAGTCCCACTACTCA 570
DB 145373 TACTAAAAATCAAAAATTAAGCCAGGTGTGTGCGACACCTGTAGTCCCACTACTCG 145314
QY 571 CGAGGTTGAGGAGAGAGATTTGCTTGAACCTAGAGAGGTGGAGTTGCAAGTAACCGAG 629
DB 145313 CGAGGCTGAGGAGAGAGATTTGCTTGAACCTGAGAGGTGGAGTTGCAAGTAACCGAGAT 145254
QY 630 --TGTCACTGACTCCAGCTCGACAGAGCAAGCACTCCATTAAGACAAAGCTTT 687
DB 145253 CGTGCACCTGCACTTCACCTCGGAGCAGAGCGAGATTCCGCCCTCAAAAAA 145194
QY 688 GAAATTTGTAAATGAGTTGTACTTCTTCTT 722
DB 145193 AAAAAGCAAAAACAAACAACTAATCTACTTT 145159

RESULT 37
AC153308 248748 bp DNA linear HTG 23-JUL-2005
LOCUS Gorilla gorilla gorilla clone CH255-200A8, WORKING DRAFT SEQUENCE,
DEFINITION 5 ordered pieces.
ACCESSION AC153308.3 GI:71067174
VERSION HTG; HTGS, PHASE2; HTGS, DRAFT.
KEYWORDS Gorilla gorilla gorilla (lowland gorilla)
SOURCE Gorilla gorilla gorilla
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Gorilla.
REFERENCE 1 (bases 1 to 248748)
Antoneilis,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,
Coleman,H., Engle,J., Franks,S., Fukenko,T., Gesteira,M.,
Greene,A., Guan,X., Gupta,T., Gurson,N., Haghighi,P., Han,J.,
Hansen,N., Ho,S.-L., Hu,P., Hunter,G., Hurtle,B., Idol,J.R.,
Kwong,P., Lalic,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R.,
Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Mastello,C.,
Mascheri,B., McDowell,J., Mojidi,H.A., Mullikin,J.C., Park,M.,
Portnoy,M.B., Prasad,A., Puri,O., Rantze,K., Reddix-Dugue,N.,
Sante,A., Schandler,K., Scheller,M.G., Sison,C., Stancirrop,S.,
Taye,A., Thomas,J.W., Thomas,P.J., Tsiouri,V., Ung,L., Vogt,J.L.,
Wetherby,K.D., Wilthers,T.R., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
JOURNAL Unpublished
TITLE 2 (bases 1 to 248748)
AUTHORS Green,E.D.
REFERENCE Direct Submission
JOURNAL Submitted (08-DEC-2004) NIH Intramural Sequencing Center, 5625
Fishers lane, Rockville, MD 20852, USA
3 (bases 1 to 248748)
Green,E.D.
REFERENCE Direct Submission
JOURNAL Submitted (23-JUL-2005) NIH Intramural Sequencing Center, 5625
Fishers lane, Rockville, MD 20852, USA
On Jul 23, 2005 this sequence version replaced gi:59676666.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nih.gov
----- Project Information
Center project name: egg
Center clone name: 200A08

```

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition,

the sequence assembly is generally based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies; the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

#### Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 246798 bases at least Q40  
 Consensus quality: 247845 bases at least Q30  
 Consensus quality: 248232 bases at least Q20  
 Insert size: 212000; agarose-ff  
 Insert size: 248348; sum-of-contigs  
 Quality coverage: 8.53x in Q20 bases; agarose-ff  
 Quality coverage: 7.28x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 55010: contig of 55010 bp in length  
 \* 55011 55110: gap of unknown length  
 \* 55111 75860: contig of 20750 bp in length  
 \* 75861 75960: gap of unknown length  
 \* 75961 94951: contig of 18991 bp in length  
 \* 94952 95051: gap of unknown length  
 \* 95052 153305: contig of 58254 bp in length  
 \* 153306 153405: gap of unknown length  
 \* 153406 248748: contig of 95343 bp in length.

#### FEATURES

##### source

1. 248748  
 /organism="Gorilla gorilla gorilla"  
 /mol\_type="genomic DNA"  
 /sub\_species="gorilla"  
 /db\_xref="taxon:9595"  
 /clone="CH255-200A8"  
 /clone\_1fb="CH255"

/note="BAC resource: <http://bacpac.chori.org/>"

misc\_feature  
 /note="clone overlaps with GenBank Accession Number AC153309 clone CH255-214B7 (center project name gtr)"

##### misc\_feature

1. 55010  
 /note="assembly\_fragment"

##### gap

vector\_end:77  
 vector\_side:left"

##### gap

55011..55110

##### misc\_feature

55111..75860

##### gap

75861..75960

##### misc\_feature

75961..94951

##### gap

94952..95051

##### misc\_feature

95052..153305

##### gap

153306..153405

##### misc\_feature

153406..248748

##### misc\_feature

/note="assembly\_fragment"

##### misc\_feature

clone\_end:SP6  
 vector\_side:right"

#### ORIGIN

Query Match 10.9%; Score 219; DB 14; Length 248748;  
 Best Local Similarity 86.9%; Pred. No. 1.6e-35;

Matches 253; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

QY 395 GGCACGGCGCAGTAGTCTGCTGTAATCCCAACACTTTGGAGGCCAAGTGGCGGA 454  
 Db 125525 GGCACGGCGCAGTAGTCTGCTGTAATCCCAACACTTTGGAGGCCAAGTGGCGGA 125584  
 QY 455 TCACCTGAGGTGAGGAGTTTAAGACACGAGCTGGCCCAACATGTGTAACCTGTCTACT 514  
 Db 125585 TTACTTGAAGTGGAGGAGTTTGAAGACATCTGCGCAACATGTGTAACCTGTCTACT 125644  
 QY 515 AAAACGCAAAATTAAGCAGGTGTGTGTGACATGACCTGTAGTCCCACTACTCAGAG 574  
 Db 125645 AAAATATCAAAATTAAGCAGGTGTGTGTGACATGACCTGTAGTCCCACTACTCAGAG 125704  
 QY 575 GTTGAAGGAGGAGGATTTGTTGAACCTTGAAGAGTGTGAGTTCAGTACCCAGAT--G 631  
 Db 125705 GTTGAAGGAGGAGGATTTGTTGAACCTTGAAGAGTGTGAGTTCAGTACCCAGATCGCG 125764  
 QY 632 TCACCTGACTCCAGCTGGCAACAGAGCAAGCTCATTAAGCAACAA 682  
 Db 125765 CACCTGTATTCCAGCTGTGTGACAGAGCAAGCTCATTAAGCAACAA 125815

#### RESULT 38

AC004079 102717 bp DNA linear PRI 27-JAN-2004  
 LOCUS Homo sapiens PAC clone RPI-167F23 from 7, complete sequence.  
 DEFINITION AC004079  
 AC004079.1 GI:2822174  
 VERSION HTG.  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

#### REFERENCE

1 (bases 1 to 102717)  
 Hillier, L.W., Fulton, R.S., Fulton, L.A., Graves, T.A., Pepin, K.H.,  
 Wagner-McPherson, C., Layman, D., Maas, J., Jaeger, S., Walker, R.,  
 Wyllie, K., Sekhon, M., Becker, M.C., O'Laughlin, M.D., Schaller, M.E.,  
 Fewell, G.A., Delahunty, K.D., Miner, T.L., Nash, W.E., Cordes, M.,  
 Du, H., Sun, H., Edwards, J., Bradshaw-Cordum, H., Ali, J., Andrews, S.,  
 Isa, A., Vanbrunt, A., Nguyen, C., Du, F., Lamer, B., Courtney, L.,  
 Kalicki, J., Ozeraky, P., Bielicki, L., Scott, K., Holmes, A.,  
 Harkins, R., Harris, A., Strong, C.W., Hou, S., Tomlinson, C.,  
 Daughin-Kohlberg, S., Kozlowski-Reilly, A., Leonard, S., Rohlfing, T.,  
 Rock, S.M., Tin-Mollam, A.M., Abbott, A., Mink, P., Maupin, R.,  
 Strommatt, C., Latreille, P., Miller, N., Johnson, D., Murray, J.,  
 Woessner, J.P., Wendt, M.C., Yang, S.P., Schultz, B.R., Mallis, J.W.,  
 Spieth, J., Bieri, T.A., Nelson, J.O., Berkowicz, N., Wohlmann, P.E.,  
 Cook, L.L., Hickendocham, M.T., Eldred, J., Williams, D., Bedell, J.A.,  
 Mardis, E.R., Clifton, S.W., Chissole, S.L., Merra, M.A., Raymond, C.,  
 Haugen, E., Gillett, W., Zhou, Y., James, R., Phelps, K., Iadonoto, S.,  
 Bubb, K., Stams, E., Levy, R., Clendinning, J., Kaul, R., Kent, W.J.,  
 Furey, T.S., Baertsch, R.A., Brent, M.R., Kibler, E., Fliscek, P.,  
 Bork, P., Suyama, M., Bailey, J.A., Portnoy, M.E., Torrents, D.,  
 Chirwalla, A.T., Gish, W.R., Eddy, S.R., McPherson, J.D., Olson, M.V.,  
 Chivler, E.E., Green, E.D., Waterston, R.H. and Wilson, R.K.

#### TITLE

The DNA sequence of human chromosome 7

#### JOURNAL

Nature 424 (6945), 157-164 (2003)

#### PUBMED

12853948

#### REFERENCE

2 (bases 1 to 102717)  
 Duckels, G., Hawkins, M., Hinds, K. and Jones, K.  
 The sequence of Homo sapiens PAC clone RPI-167F23

#### JOURNAL

Unpublished (2001)

#### REFERENCE

3 (bases 1 to 102717)  
 Waterston, R.  
 Direct Submission

#### JOURNAL

Submitted (29-JAN-1998) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

#### REFERENCE

4 (bases 1 to 102717)  
 Waterston, R.  
 Direct Submission

#### JOURNAL

Submitted (03-FEB-2000) Department of Genetics, Washington

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
5 (bases 1 to 102717)  
Waterston, R.  
Direct Submission  
Submitted (04-FEB-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 102717)  
Waterston, R.  
Direct Submission  
Submitted (26-APR-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
7 (bases 1 to 102717)  
Wilson, R.  
Direct Submission  
Submitted (27-JAN-2004) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_DJ0167F23  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
This clone was provided for sequencing by Dr. Stephen Scherer,  
Department of Genetics, The Hospital for Sick Children, Toronto,  
Ontario, Canada, with support from the Canadian Genome Analysis and  
Technology Program; and Dr. John D. McPherson, Department of  
Genetics, Washington University, St. Louis MO. For additional  
information about the map position of this sequence, see  
<http://www.genet.sickkids.on.ca/chromosome7> and  
<http://genome.wustl.edu>

SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-1, prepared by  
Pieter de Jong and coworkers at <http://www.chori.org> using the  
method described by Ioannou et al., Nature Genetics 6:84-9 (1994).  
The library is from one male donor.  
The clone may be obtained either from Genome Systems, Inc.  
(<http://www.genomesystems.com>) or Research Genetics, Inc.  
(<http://www.resgen.com>); or from Pieter de Jong.  
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:  
The actual start of this clone is at base position 1 of RP1-167F23  
the actual end is at base position 102717 of RP1-167F23. The  
orientation of this clone is unknown.

FEATURES  
source  
Location/Qualifiers  
1..102717  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7"  
/clone="RP1-167F23"  
/clone\_id="RPCI-1"  
1..1098  
/rpt\_family="L1"  
repeat\_region  
/rpt\_family="L1"

repeat\_region 1099..1412  
/rpt\_family="Alu"  
repeat\_region 1413..2186  
/rpt\_family="L1"  
repeat\_region 2187..2495  
/rpt\_family="Alu"  
repeat\_region 2496..2910  
/rpt\_family="L1"  
repeat\_region 3131..3332  
/rpt\_family="Alu"  
repeat\_region 3333..3696  
/rpt\_family="Mair"  
repeat\_region 3697..3799  
/rpt\_family="Alu"  
repeat\_region 4056..4081  
/rpt\_family="(TTTC)n"  
repeat\_region 4145..4277  
/rpt\_family="L2"  
repeat\_region 4926..5064  
/rpt\_family="MER1\_type"  
repeat\_region 5081..5332  
/rpt\_family="L1"  
repeat\_region 6427..6546  
/rpt\_family="Alu"  
repeat\_region 6572..6868  
/rpt\_family="Alu"  
repeat\_region 7942..8236  
/rpt\_family="Alu"  
repeat\_region 8544..8579  
/rpt\_family="AT-rich"  
repeat\_region 8991..9300  
/rpt\_family="Alu"  
repeat\_region 10140..10406  
/rpt\_family="L1"  
repeat\_region 10407..10702  
/rpt\_family="Alu"  
repeat\_region 10703..10790  
/rpt\_family="L1"  
repeat\_region 11252..11503  
/rpt\_family="MIR"  
repeat\_region 11814..11854  
/rpt\_family="MIR"  
repeat\_region 12097..12443  
/rpt\_family="Alu"  
repeat\_region 12503..12554  
/rpt\_family="MIR"  
repeat\_region 12555..12687  
/rpt\_family="Alu"  
repeat\_region 12705..12783  
/rpt\_family="MIR"  
repeat\_region 13560..13859  
/rpt\_family="Alu"  
repeat\_region 13863..13976  
/rpt\_family="MIR"  
repeat\_region 14007..14163  
/rpt\_family="MIR"  
repeat\_region 14385..14604  
/rpt\_family="MIR"  
repeat\_region 15123..15178  
/rpt\_family="L2"  
repeat\_region 15754..16057  
/rpt\_family="Alu"  
repeat\_region 16104..16126  
/rpt\_family="AT-rich"  
repeat\_region 16134..16443  
/rpt\_family="Alu"  
repeat\_region 16868..16975  
/rpt\_family="MIR"  
repeat\_region 16995..17352  
/rpt\_family="Mair"  
repeat\_region 17356..17434  
/rpt\_family="L1"  
repeat\_region 17435..17736

```

repeat_region      /rpt_family="Alu"
                    /rpt_family="Alu"
repeat_region      17737..18018
                    /rpt_family="Alu"
repeat_region      18015..18495
                    /rpt_family="Alu"
repeat_region      18498..18788

Query Match
Best Local Similarity 61.4%; Pred. No. 2.1e-35; Length 102717;
Matches 421; Conservative 0; Mismatches 257; Indels 8; Gaps 4;

QY 1 TSATCCACGAGCTTGGCTCCCAAGTCTGGGATTAAGGCGTGAAGCCACCCGCTG 60
Db 52054 TATTCGCGCCGCTGGCTCCCTAGTGTGGGATTAAGGCGTGAAGCCACCCG 52113
QY 61 GTCGAATGCTTTATTTATTTGAAGACACATGGGCTTAAATCTGCTTCTATTGAC 120
Db 52114 GCC-AGGGTTGAATTTTAAATGACATTTTACATTTGAGAAAGATTGTACCAAAATC 52171
QY 121 AGACTTGTAGTGAATCCCAATGCGCCACTTACGAAAGGCTTAAATGACTTAC 180
Db 52172 ATTCATTTGCATTTTACGAAATGACATTAACATTTTATGCTATTTATTAAGCTGA 52231
QY 181 TCTCTCTCAGCTGTCTTCTGCAATGTAAGGTGAATATGATGCTTTCAAGAGAA 240
Db 52232 AGTATTAAGAAAGTGTATTATTTTATTTAGTATCTTCTTCAATTTTGAAGAAC 52291
QY 241 TAAACCTATGAAAAGTGTGAGAGATGTTGATGATGAATTAAGATTCAACAAGTAC 300
Db 52292 TAGAGAAAAGTGTACAAAAGAAACACCCCAAAATCTTACTTAACGAATTA 52351
QY 301 TACCTGCTATGAAAGTTTAAGAGTTATTTATTAACAATTTTAAATTTTAAAC 360
Db 52352 AAGCTATGAAAGTGTGAGCTGCTATTCATTAATGTTCTTAAATGTTATTCCTTAT 52411
QY 361 TATACACTTAATTTATTAAGAGCTTTGAAATGGGCGAGGAGTACCTGCTGCTG 420
Db 52412 TATATACCTGCTTTATTTAAATTAAGT--GTCTCGGCGGAGGAGTGTGCTGACACCT 52469
QY 421 AATCCCAACACTTTGGGAGGCGCAAGTGGGCGGATCCTGAGGTCAGAGATTTAAGAC 480
Db 52470 AATCCCAACACTTTGGGAGGCGTGAAGTGGGCGGATCCTGAGGTCAGAGATTTAAGAC 52529
QY 481 AGCTGCGCAACATGTTGAACCCCTGTCTCTACTTAAACCGCAAAATTAAGCAGGTG 540
Db 52530 AGCTGCGCAACATGTTGAACCCCTGTCTCTACTTAAACCGCAAAATTAAGCAGGTG 52589
QY 541 GTGGAGTGAACCTGTGATGCTCCCACTACGAGAGTGTGAGGAGGAGATTGCTGAAC 600
Db 52590 GTGGAGTGAACCTGTGATGCTCCCACTACGAGAGTGTGAGGAGGAGATTGCTGAAC 52649
QY 601 TAGAGAGTGAAGTGTGACATTAACCCGAGAT--GTCACTGCACTCCAGCT--GGCAACAG 656
Db 52650 CAGGAGGTGAGAGTGTGACATTAACCCGAGAT--GTCACTGCACTCCAGCT--GGCAACAG 52709
QY 657 AGCAAGACTCCATTAAGACACAAA 682
Db 52710 ACCGAGACTCCGCTTAACAAA 52735

RESULT 39
AC011380 133955 bp DNA linear PRI 01-NOV-2002
LOCUS Homo sapiens chromosome 5 clone CTB-133C10, complete sequence.
AC011380
VERSION AC011380.6 GI:24462307
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE 1 (bases 1 to 133955)

```

```

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 133955)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 133955)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
TITLE Submitted (01-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Nov 1, 2002 this sequence version replaced gi:8576057.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source
location/Qualifiers
1..133955
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-133C10"

ORIGIN
Query Match
Best Local Similarity 77.4%; Pred. No. 2e-35; Length 133955;
Matches 291; Conservative 0; Mismatches 82; Indels 3; Gaps 2;

QY 372 AATTATTAAGAGCTTTGAATGAGCGCAGGCGAGTAGTCTGCTGATATCCAAAC 431
Db 26185 AATCTTTAAAGTGTCTTATTAATGCGCAGGCGAGTAGTCTGCTGATATCCAAAC 26244
QY 432 TTTGGAGCGCAAGTGGCGGATGATCCTGAGGTGAGAGTTTAAGCAGGCTGCCAA 491
Db 26245 TTTGGAGCGCAAGTGGCGGATGATCCTGAGGTGAGAGTTTAAGCAGGCTGCCAA 26304
QY 492 CATGTTGAACCTGTGCTCTACTTAAACCGCAAAATTAAGCAGGCGGAGTGGCGAC 551
Db 26305 CATGTTGAACCTGTGCTCTACTTAAACCGCAAAATTAAGCAGGCGGAGTGGCGAC 26364
QY 552 CTGTAGTCCCACTACTGAGAGTGTGAGGAGGAAATTTGCTTAACCTAGAGGTGA 611
Db 26365 CTGTAGTCCCACTACTGAGAGTGTGAGGAGGAAATTTGCTTAACCTAGAGGTGA 26424
QY 612 GGTTCAGTAAACCGAGATGTCAC--TGCACTCCAGCTGGCAACAGCAAGACTCCAT 669
Db 26425 GGTTCAGTAAACCGAGATGTCAC--TGCACTCCAGCTGGCAACAGCAAGACTCCAT 26484
QY 670 -AAAGACAAACAACTTTGAATGTTGAATGAGTTGTACTTATCTTATTAAGAA 728
Db 26485 CTCAAAAAAAGTCTTATTTGTGTGTCACATGCTCTCTAAGTCTTATTAACA 26544
QY 729 TTCACTTTTGTTCATT 744
Db 26545 TTCACTTTTATTCCTT 26560

RESULT 40
AC079931 178459 bp DNA linear HTG 16-MAR-2001
LOCUS Homo sapiens chromosome 7 clone CTD-2594L23, WORKING DRAFT
AC079931
DEFINITION SEQUENCE, 43 unordered pieces.
AC079931.3 GI:13357557
VERSION AC079931.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
AUTHORS 1 (bases 1 to 178459)  
TITLE Waterston.R.H.  
JOURNAL The sequence of Homo sapiens clone  
AUTHORS 2 (bases 1 to 178459)  
TITLE Waterston.R.H.  
JOURNAL Unpublished  
TITLE Direct Submission  
JOURNAL Submitted (17-SEP-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Mar 16, 2001 this sequence version replaced gi:11991333.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_T02594L23  
----- Summary Statistics -----  
Sequencing vector: MJ3, 95%  
Sequencing vector: plasmid, 5%  
Chemistry: Dye-terminator Big Dye, 5% of reads  
Chemistry: Dye-terminator Big Dye, 5% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 154148 bases at least Q40  
Consensus quality: 162306 bases at least Q30  
Consensus quality: 166493 bases at least Q20  
Insert size: 191000; agarose-fp  
Insert size: 174259; sum-of-contigs  
Quality coverage: 3.14 in Q20 bases; agarose-fp  
Quality coverage: 3.53 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 43 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 1135: contig of 1135 bp in length  
\* 1136 1135: gap of unknown length  
\* 1235: gap of unknown length  
\* 1236 2532: contig of 1297 bp in length  
\* 2533 2632: gap of unknown length  
\* 2633 3791: contig of 1159 bp in length  
\* 3791 3891: gap of unknown length  
\* 3891 5003: contig of 1112 bp in length  
\* 5003 5103: gap of unknown length  
\* 5103 6609: contig of 1506 bp in length  
\* 6609 6709: gap of unknown length  
\* 6709 8108: contig of 1399 bp in length  
\* 8108 8209: gap of unknown length  
\* 8209 9889: contig of 1681 bp in length  
\* 9889 9989: gap of unknown length  
\* 9989 12104: contig of 2115 bp in length  
\* 12104 12204: gap of unknown length  
\* 12204 14028: contig of 1824 bp in length  
\* 14028 14128: gap of unknown length  
\* 14128 15404: contig of 1276 bp in length  
\* 15404 15504: gap of unknown length  
\* 15504 17820: contig of 2316 bp in length  
\* 17820 17920: gap of unknown length  
\* 17920 20527: contig of 2607 bp in length  
\* 20527 20628: gap of unknown length  
\* 20628 23434: contig of 2807 bp in length  
\* 23434 23534: gap of unknown length  
\* 23534 26463: contig of 2929 bp in length  
\* 26463 26564: gap of unknown length  
\* 26564 28733: contig of 2170 bp in length  
\* 28733 28833: gap of unknown length  
\* 28833 30672: contig of 1839 bp in length  
\* 30672 30772: gap of unknown length

\* 30772 30772: contig of 2622 bp in length  
\* 30772 33395: gap of unknown length  
\* 33395 33494: gap of unknown length  
\* 33494 35974: contig of 2480 bp in length  
\* 35974 35975: gap of unknown length  
\* 35975 36074: gap of unknown length  
\* 36074 38223: contig of 2149 bp in length  
\* 38223 38324: gap of unknown length  
\* 38324 40744: contig of 2421 bp in length  
\* 40744 40844: gap of unknown length  
\* 40844 44797: contig of 3953 bp in length  
\* 44797 44897: gap of unknown length  
\* 44897 48446: contig of 3549 bp in length  
\* 48446 48546: gap of unknown length  
\* 48546 48547: gap of unknown length  
\* 48547 52596: contig of 4050 bp in length  
\* 52596 52696: gap of unknown length  
\* 52696 57018: contig of 4322 bp in length  
\* 57018 57118: gap of unknown length  
\* 57118 60603: contig of 3485 bp in length  
\* 60603 60703: gap of unknown length  
\* 60703 60704: gap of unknown length  
\* 60704 62021: contig of 1318 bp in length  
\* 62021 62121: gap of unknown length  
\* 62121 67532: contig of 5411 bp in length  
\* 67532 67632: gap of unknown length  
\* 67632 72176: contig of 4544 bp in length  
\* 72176 72276: gap of unknown length  
\* 72276 76066: contig of 3790 bp in length  
\* 76066 76166: gap of unknown length  
\* 76166 80719: contig of 4553 bp in length  
\* 80719 80819: gap of unknown length  
\* 80819 84399: contig of 3580 bp in length  
\* 84399 84400: gap of unknown length  
\* 84400 84499: gap of unknown length  
\* 84499 88673: contig of 4173 bp in length  
\* 88673 88772: gap of unknown length  
\* 88772 95190: contig of 6418 bp in length  
\* 95190 95290: gap of unknown length  
\* 95290 95291: gap of unknown length  
\* 95291 100594: contig of 5304 bp in length  
\* 100594 100595: gap of unknown length  
\* 100595 106783: contig of 6089 bp in length  
\* 106783 106883: gap of unknown length  
\* 106883 113874: contig of 6991 bp in length  
\* 113874 113974: gap of unknown length  
\* 113974 118494: contig of 4520 bp in length  
\* 118494 118594: gap of unknown length  
\* 118594 126676: contig of 8082 bp in length  
\* 126676 126777: gap of unknown length  
\* 126777 134408: contig of 7632 bp in length  
\* 134408 134508: gap of unknown length  
\* 134508 145575: contig of 11067 bp in length  
\* 145575 145576: gap of unknown length  
\* 145576 153665: contig of 7890 bp in length  
\* 153665 153666: gap of unknown length  
\* 153666 163958: contig of 10293 bp in length  
\* 163958 164058: gap of unknown length  
\* 164058 178459: contig of 14401 bp in length.  
\* 178459 Location/Qualifiers

## FEATURES

## source

1. 178459

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="7"

/clone="CTD-2594L23"

1. 1135

/note="assembly\_name: Contig32"

1136. 1235

/estimated\_length=unknown

1236. 2532

/note="assembly\_name: Contig33"

2533. 2632

/estimated\_length=unknown

2633. 3791

/note="assembly\_name: Contig34"

3792. 3891

/estimated\_length=unknown

## misc\_feature

## gap

## misc\_feature

## gap

## misc\_feature

## gap

```

misc_feature      3892..5003
                    /note="assembly_name:Contig37"
gap                5004..5103
                    /estimated_length=unknown
misc_feature      5104..6609
                    /note="assembly_name:Contig38"
gap                6610..6709
                    /estimated_length=unknown
misc_feature      6710..8108
                    /note="assembly_name:Contig39"
gap                8109..8208
                    /estimated_length=unknown
misc_feature      8209..9889
                    /note="assembly_name:Contig43"
gap                9890..9989
                    /estimated_length=unknown
misc_feature      9990..12104
                    /note="assembly_name:Contig46"
gap                12105..12204
                    /estimated_length=unknown
misc_feature      12205..14028
                    /note="assembly_name:Contig47"
gap                14029..14128
                    /estimated_length=unknown
misc_feature      14129..15404
                    /note="assembly_name:Contig49"
gap                15405..15504
                    /estimated_length=unknown
misc_feature      15505..17820
                    /note="assembly_name:Contig50"
gap                17821..17920
                    /estimated_length=unknown
misc_feature      17921..20527
                    /note="assembly_name:Contig51"
gap                20528..20627
                    /estimated_length=unknown
misc_feature      20628..23434
                    /note="assembly_name:Contig52"
gap                23435..23534
                    /estimated_length=unknown
misc_feature      23535..26463
                    /note="assembly_name:Contig53"
gap                26464..26563
                    /estimated_length=unknown
misc_feature      26564..28733
                    /estimated_length=unknown

Query Match      10.9%; Score 218.8; DB 14; Length 178459;
Best Local Similarity 61.4%; Pred. No. 1.9e-35;
Matches 421; Conservative 0; Mismatches 257; Indels 8; Gaps 4;

QY      1 TGAATCAGCAGCTTGGCTCCCAAGTGTGGATTACAGGCGTGAAGCCACGCGCTG 60
        |||||
DB      68469 TGAATCGCGCGCTCGCGCTCCCTTAAGTGTGGATTACAGGCGTGAAGCCACGCGCA 68528

QY      61 GTGGAATGCTTATTTGAAGACAAATGCGCTTAATCTGCTCATTTGAC 120
        |||||
DB      68529 GCC--AGGGTTGAATATTTTAAATGCACTTTACCTTGAGAAATGTGACAAATC 68586

QY      121 AGACTTTGATGAGTCAATCCCAATGCTGCACTTGAACGCGCTTAATGACTT 180
        |||||
DB      68587 ATCCATTGCACTTTGAGTAATGACATTAATTAAGTCACTTATATTAAGTGA 68646

QY      181 TCTCTCGAGCTGCTTTCTGCAATGTAAAGTGAATATGATGCGCTTCAAGAGGAA 240
        |||||
DB      68647 AGTATTAAGAAATGTTGTTATATTTTAAATGATCTTAATTTGAGAGAACCC 68706

QY      241 TAAACTATGAAGAAAGTGTGAGATGTGTTGATATGAATTAAGATTTCAACAATG 300
        |||||
DB      68707 TAGAGAAAAAGTTTACAAAAAGAAAAACACCCCAAAATCTACTTTAAAGATTA 68766

QY      301 TAGCTGCTATTGAAGATTATTAAGATTATTAACAATTAATTAATTAATTAATTA 360
        |||||
DB      68767 AAGCTAATGAAGTAGGCTGCTATTTATTTAGTTCCCTAATATGCTCATTTCCCTTA 68826

```

```

QY      361 TATACACTTAATTTATTAAGAGCTTTGAAATGAGCCAGCGCAGTAGCTCTGCTGT 420
        |||||
DB      68827 TATTAATCTGCTTTTAAATAACT--GTCTCGGCCCGGAGATGTGTGCTGACACTGT 68884

QY      421 AATCCCAACACTTTGGAGGCCCAAGGTGGCGGATCACTGATGAGAGATTAAAGCC 480
        |||||
DB      68885 AATCCAGACACTTTGGAGGGCTGAGGTGGATGATCACTGAGAGTGGGAGTTCAAGAGC 68944

QY      481 AGCTTGCCCAACATGATGAACCTTGTCTTACTATAAAACGCAAAATTAAGCAGGTGTG 540
        |||||
DB      68945 AGCTTGCCCAACATGATGAACCTTGTCTTACTATAAAACGCAAAATTAAGCAGGTGTG 69004

QY      541 GTGGCATGACCGTGTATGCCCACTACTCAGAGAGTTGAGGAGAGAAATTTGCTTGACC 600
        |||||
DB      69005 GTGGTGGGCACTGTATATTCAGACTTCTGGAGGCTGAGAGGAGGAAATTTGTTAACC 69064

QY      601 TAGAGGTGAGAGTTGACAGTAACCCGAGAT--GTCACTGCACTCCAGCT--GGCAACAG 656
        |||||
DB      69065 CAGAGGTGAGAGTTGACAGTAACCCGAGAT--GTCACTGCACTCCAGCT--GGCAACAG 69124

QY      657 AGCAAGACTTCATTAAGACAAACAAA 682
        |||||
DB      69125 ACCGAGACTCCGCTTAATAAAAAAAA 69150

RESULT 41
AL136365/c
LOCUS      AL136365      179006 bp      DNA      linear      PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-143M15 on chromosome 9p23-24.3
              contains the DMR1 gene for doublesex and mab-3 related
              transcription factor 1, the DMR3 gene for doublesex and mab-3
              related transcription factor 3 and four CpG islands, complete
              sequence.
ACCESSION   AL136365
VERSION     AL136365.9  GI:8246868
KEYWORDS    HTG; CpG island; DMR1; DMR3; doublesex; mab-3.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE   1 (bases 1 to 179006)
AUTHORS    Laird G.
TITLE       Direct Submision
JOURNAL     Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT     On Jun 4, 2000 this sequence version replaced GI:7801526.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            -----
            Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vegas@sanger.ac.uk
            -----
            RP11-143M15 is from the library RBC1-11.1 constructed by the group
            of Pieker de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBAC3.6
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least

```





QY 239 AATTAACCTATGAAAGTGTGAGATAGTGTGATATGAATAAGATTCAACAAGT 298  
 175142 TGGATACAGTAAGATATCTTAACTAGTGCATTAACATATGTTCTTGAAGATT 175083  
 QY 299 AGTAGCTGCTATTGAAGATTGAAGATTATTAACAATTATTAATAATTTTAAAA 358  
 175082 CAGGATCTGACAGTAATCTCAATATTAATTAAGATATGACAGACAGCTGGCTC 175023  
 QY 359 ACTAATACACTTAATTAATTAAGAGCTTTGAATGAGCCGAGAGTATGCTCCCT 418  
 175022 ACAGAGCTATCTTGAAGATGAATAACTGATTAATGCGTGGTGAAGTCTCAATACCT 174963  
 QY 419 GTTAATCCCAACTTTTGGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAG 478  
 174962 GTTAATCCCAACTTTTGGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAG 174905  
 QY 479 CCAGCTGGCCAACTGATGAACCTGTCTCTAATAAAGCCAAATTAAGCCAGTGG 538  
 174904 CCATCTGGCCAACTGATGAACCTGTCTCTAATAAAGCCAAATTAAGCCAGTGG 174845  
 QY 539 TGGTGCATGCACTCTGATGCTCCCACTACTCAGAGAGTTGAGGAGAGAAATGCTTGA 598  
 174844 TGGTGCATGCACTCTGATGCTCCCACTACTCAGAGAGTTGAGGAGAGAAATGCTTGA 174785  
 QY 599 CCTAGGAGGTGAGAGTTGAGAGTCAACCCGAGATGTCAC---TGCACTCCAGCTGGCAACA 655  
 174784 CCGGAGAGGTGAGAGTTGAGAGTCAACCCGAGATGTCACCTATTGACCTCCAGCTGGTGA 174725  
 QY 656 GAGCAAGACTCCAT-----AAGACACACAAAGCTTTGAATTTGTAATGATGTTGA 709  
 174724 GAGTGAAGCTCCATGTCACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 174665  
 QY 710 CCTATCTTGA-TTTAAGAAATTCATCTTTGTTCAATTTATTTTACT 754  
 174664 TCTGTGGGCAATTTTCATTTCAATTTGTTGTTTAAATGATGCT 174619  
 Db  
 RESULT 42  
 HSDPA/c 305000 bp DNA linear PRI 29-MAY-2003  
 LOCUS Homo sapiens chromosome X sequence from 6 PACs 1 BAC and 1 cosmid,  
 region GJB1-DXS559 map Xq13.1; segment 1/2, complete sequence.  
 ACCESSION AL590762.1 GI:13751769  
 VERSION AL590762.1 GI:13751769  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiinae; Homo.  
 REFERENCE  
 AUTHORS 1  
 Ramer, J., Heilmann, K., Sudbrak, R., Kosiura, A., Klages, S.,  
 Steffens, C., Borzym, K., Kube, M., Lehrach, S., Marguardt, I.,  
 Schuelch, S., Starke, A., Thompson, C., Hennig, S., Francis, F.,  
 Nemeth, A., Monaco, A., Lehrach, H. and Reinhardt, R.  
 Unpublished  
 2 (bases 1 to 305000)  
 JOURNAL Direct Submission  
 TITLE Submitted (15-FEB-1999) MPMG, Abt. Lehrach, Max Planck Institut  
 AUTHORS Puert Molekulare Genetik, Imnestrasse 73, Berlin, 14195 Germany  
 COMMENT On May 14, 2001 this sequence version replaced gi:11228442.  
 Baes 300001..305000 overlap with AL590763 (HSDPB).  
 FEATURES  
 source  
 1..305000  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /map="Xq13.1"  
 /clone="PAC RPCI-1 56F19"  
 /clone\_1ib="RPCI1.3-5 Human PAC library, originating

ORIGIN  
 Query Match 10.9%; Score 218.8; DB 8; Length 305000;  
 Best Local Similarity 62.0%; Pred. No. 1.7e-35;  
 Matches 424; Conservative 0; Mismatches 237; Indels 23; Gaps 4;  
 /note="region between markers GJB1-DXS559"  
 QY 1 TGAATCCAGAGCTTGGCTCCCAAGTGTGATTAAGAGGCTGAGCCACCGCTG 60  
 117592 TGAATCCAGAGCTTGGCTCCCAAGTGTGATTAAGAGGCTGAGCCACCGCTG 117533  
 QY 61 GTGCAATGCTTATTAATTAATTAAGAGCAATGAGCTTAAATCTGCTTCAATTTGAC 120  
 117532 GAC-TTAAATCTGTTTGAAGAAATCAATTAATTAATCTTCAATTTGAGCA 117474  
 QY 121 AGACTTGTGATGATCAATTAATCCCAATGCTGCACTTAAGTAAACGGCTTAAATGACTTAG 180  
 117473 ATGAATTTCTTAATTAATTAATCAATCCCAATGCAATTAATTAATTAATTAATTTG 117414  
 QY 181 TCTCTCTGCTGCTCTTCTTGCATATGTAAGTGAATATGATGCTTTCAAGAGGAA 240  
 117413 ACCACATCA-----AATGATAAATTTTGGGGGGGGTGTCTTAAATAAAT 117369  
 QY 241 TAAACCTATGAAGAAAGTGTGAGATAGTGTGATTAAGATTTCAACAGTAG 300  
 117368 TAAATCTTTTA-----TCTTCAATTAATTAATTAAGAAATTAATTAATTAATTAAT 117313  
 QY 301 TAGCTGCTATTAAGATTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 360  
 117312 TGGGAAAAATTTTGTAAATATCAATCATCTGATTAAGGGCTGTATCCAGATATACAA 117253  
 QY 361 TAATACCTTAATTAATTAAGAGCTTTGAATGAGGCGAGCTAGCTCTGCTGT 420  
 117252 GAATCTTCAACTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 117193  
 QY 421 AATCCCAACTTTTGGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGAC 480  
 117192 AATCCCAACTTTTGGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGAC 117133  
 QY 481 AGCTGGCCCAATGATGTAACCTGTCTCTACTTAAAGCAAAATTAAGCAGGTGTG 540  
 117132 AGCTGGCCCAATGATGTAACCTGTCTCTACTTAAAGCAAAATTAAGCAGGTGTG 117073  
 QY 541 GTGGATGCACTGTGATGCTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 600  
 117072 GTGGATGCACTGTGATGCTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 117013  
 QY 601 TGAAGAGTGAAGTTGCAATACCCGAGT---GTCACTGCACTCAGCTTGGCAACAG 657  
 117012 TGAAGAGTGAAGTTGCGGTGAGCCAGATTTGCGCATTTGCACTCAGCTTGGTGA 116953  
 QY 658 GCAAGCTCCATTAAGACAAACAA 681  
 116952 GTGAATCTCATCTCAAAAAAAA 116929  
 Db  
 RESULT 43  
 HS568C11  
 LOCUS Human DNA sequence from clone RP4-568C11 on chromosome  
 DEFINITION 20p11.21-11.23 Contains the CST7 gene for cystatin F  
 (leukocystatin) (cystatin 7, leukocystatin, cystatin-1 like  
 metastasis-associated protein, CMAP), the gene C20ORF3 for  
 chromosome 20 open reading frame 3, the 3' end of the ACAS2L gene  
 for acetyl-Coenzyme A synthetase (AMP forming)-like and a CpG  
 island, complete sequence.  
 ACCESSION AL035661.16 GI:6015535  
 VERSION HTG; ACAS2L; AMP-binding; C20ORF3; CMAP; CpG island; CST7;  
 cystatin; leukocystatin.  
 KEYWORDS  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
1 (bases 1 to 103681)  
Barlow,K.  
Direct Submission  
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Oct 7, 1999 this sequence version replaced gi:6002138.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20  
RP4-568C11 is from the library RPCT-4 constructed by the group of  
Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.

Location/Qualifiers  
1. 103681  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="p11.21-11.23"  
/clone="RP4-568C11"  
/clone\_1fb="RPCT-4"  
1  
/note="Clone left\_end: RP4-568C11"  
25136..35834  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
join(25136..25481,33193..33365,34834..34950,35541..35834)  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
/note="match: ESTs: AA089317 AA089339 A1386371 A1707662  
A1734965 A1927615 AL559378 AL559379 AM467354 AM514304  
BF686825 BF523613 BF898503 BG515774 BG547285 B1768143  
B1837059 BM253995 BM363008 BM36306 BM363506 BM363782  
BM363821 BM365975 BM561663 BM564264 BM564523 BM917046  
BQ051538 BQ051619 BQ052118 BQ053137 BQ053396 BQ053551  
BQ054014 BQ054230 N47763  
match: CDNAS: AB015225 AF031824 AF036342 BC015507  
join(25346..25481,33193..33365,34834..34950,35541..35618)  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
/standard\_name="OTTHUMP0000030465"  
/note="match: proteins: O76096 QUBED4"  
/codon\_start=1  
/protein\_id="CAB75498.2"  
/db\_xref="GI:56203089"  
/db\_xref="Genew:2479"  
/db\_xref="GOA:O76096"  
/db\_xref="HSSP:IG96"

FEATURES  
source

misc\_feature  
1  
/note="Clone left\_end: RP4-568C11"  
25136..35834  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
join(25136..25481,33193..33365,34834..34950,35541..35834)  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
/note="match: ESTs: AA089317 AA089339 A1386371 A1707662  
A1734965 A1927615 AL559378 AL559379 AM467354 AM514304  
BF686825 BF523613 BF898503 BG515774 BG547285 B1768143  
B1837059 BM253995 BM363008 BM36306 BM363506 BM363782  
BM363821 BM365975 BM561663 BM564264 BM564523 BM917046  
BQ051538 BQ051619 BQ052118 BQ053137 BQ053396 BQ053551  
BQ054014 BQ054230 N47763  
match: CDNAS: AB015225 AF031824 AF036342 BC015507  
join(25346..25481,33193..33365,34834..34950,35541..35618)  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
/standard\_name="OTTHUMP0000030465"  
/note="match: proteins: O76096 QUBED4"  
/codon\_start=1  
/protein\_id="CAB75498.2"  
/db\_xref="GI:56203089"  
/db\_xref="Genew:2479"  
/db\_xref="GOA:O76096"  
/db\_xref="HSSP:IG96"

misc\_feature  
32890..33079  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
/note="Random repeat. Region contains single clone and  
forced join. Assembly is consistent with the digest."  
35803..35808  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
35834  
/gene="CST7"  
/locus\_tag="RP4-568C11.3-001"  
complement(38831..68657)  
/locus\_tag="RP4-568C11.2-001"  
complement(join(38831..39928,44798..44990,45432..45566,  
46103..46277,47366..47482,49551..49643,54673..54788,  
59803..59925,68500..68657))  
/gene="C20orf3"  
/locus\_tag="RP4-568C11.2-001"  
/note="match: ESTs: AA023064 AA057573 AA078933 AA079231  
AA128202 AA130982 AA164222 AA172232 AA179409 AA179549  
AA190882 AA193559 AA199685 AA256230 AA287697 AA295645  
AA295719 AA299550 AA299890 AA308433 AA335036 AA345847  
AA350163 AA374277 AA393476 AA455403 AA464128 AA478826  
AA479994 AA552922 AA602522 AA604095 AA632605 AA709451  
AA737976 AA774045 AA785837 AA810442 AA853299  
AA861665 AA892916 AA908661 AA932982 AA974843 AA975875  
AA989411 AT032386 AT042188 AT125617 AT140761 AT149766  
AT200324 AT205248 AT208259 AT278121 AT301561 AT345846  
AT470112 AT500350 AT526385 AT560081 AT750684 AT750995  
AT750996 AT754666 AT754907 AT811737 AT819701 AT826600  
AT828005 AT884917 AT961461 AT991422 AM005061 AM023303  
AA028593 D20431 F12233 H12659 H26812 H26824 H27474 H28147  
H47768 H82670 H94409 N32314 N40320 N41442 N57409 R11259  
R12093 R13491 R18192 R37399 R42258 R54498 R55919 R56192  
R71351 R79866 R83729 T03538 T23653 T23663 T30196  
T34150 T34847 T48031 T55377 T55805 T66394 T73292 T73306  
T73353 T73370 T77966 T84396 T93282 W63689 Z19171 Z28429  
Z43384 Z44002  
match: CDNAS: M57547"  
complement(38831)  
/gene="C20orf3"  
/locus\_tag="RP4-568C11.2-001"  
complement(38850)  
/gene="C20orf3"  
/locus\_tag="RP4-568C11.2-001"  
complement(38864..38869)  
/gene="C20orf3"  
/locus\_tag="RP4-568C11.2-001"  
complement(join(39719..39928,44798..44990,45432..45566,  
46103..46277,47366..47482,49551..49643,54673..54788,  
59803..59925,68500..68657))  
/gene="C20orf3"  
/locus\_tag="RP4-568C11.2-001"  
/standard\_name="OTTHUMP0000030464"  
/note="match: proteins: CE06440 CE16156 O81484 P15324  
P18417 Q24160 Q9ZVB7"  
/codon\_start=1  
/protein\_id="CAB75499.1"  
/db\_xref="GI:6996428"  
/db\_xref="Genew:13238"  
/db\_xref="Genew:13238"  
/db\_xref="InterPro:IPR004141"  
/db\_xref="UniProt/Swiss-Prot:O9HDC9"  
/translation="MSADGKQRPLRQVVTDDGQPEAKGSSFGGRVRRVTL  
MAVSLVPLGAMMLBSPIDPPLSFKEPRLLLGVLPHTKLRQAEKRLFENQLVGP



REFERENCE 5 (bases 1 to 181174)  
 AUTHORS Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Mar 1, 2002 this sequence version replaced gi:18652479.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT

FEATURES Location/Qualifiers

source

1..181174  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /clone="RP11-9N20"

misc\_feature

complement(1..2004)  
 /note="overlaps bases 1..2004 of clone AC117381"

repeat\_region

complement(188..560)  
 /function="clone overlap"

repeat\_region

3310..3337  
 /rpt\_family="MT1A1"

repeat\_region

5323..5377  
 /rpt\_family="T1MA5A"

repeat\_region

complement(5378..5389)  
 /rpt\_family="T1MA5A"

repeat\_region

complement(5386..5853)  
 /rpt\_family="T1M2"

repeat\_region

complement(5846..6339)  
 /rpt\_family="T1M2"

repeat\_region

complement(7364..7707)  
 /rpt\_family="T1M2"

repeat\_region

complement(7904..8205)  
 /rpt\_family="AluYo"

repeat\_region complement(8597..8682)  
 /rpt\_family="L2"  
 repeat\_region 8683..8990  
 /rpt\_family="AluSg"  
 repeat\_region complement(8991..9981)  
 /rpt\_family="L2"  
 repeat\_region complement(10829..11216)  
 /rpt\_family="MSTD"  
 repeat\_region 11899..11925  
 /rpt\_family="AT\_rich"  
 repeat\_region 12045..12346  
 /rpt\_family="AluSx"  
 repeat\_region 12356..12460  
 /rpt\_family="GGAAn"  
 repeat\_region 13263..13284  
 /rpt\_family="(A)n"  
 repeat\_region complement(13499..13940)  
 /rpt\_family="LTR40b"  
 repeat\_region complement(13966..14096)  
 /rpt\_family="MIR"  
 repeat\_region complement(14179..14317)  
 /rpt\_family="MIR"  
 repeat\_region complement(14366..14643)  
 /rpt\_family="AluJb"  
 repeat\_region complement(16681..16824)  
 /rpt\_family="MIR"  
 repeat\_region 17093..17211  
 /rpt\_family="L2"  
 repeat\_region 17980..18260  
 /rpt\_family="AluSx"  
 repeat\_region 18261..18288  
 /rpt\_family="(GAAA)n"  
 repeat\_region 18471..18501  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(19095..19176)  
 /rpt\_family="MER94"  
 repeat\_region 19264..19555  
 /rpt\_family="AluSx"  
 STS 20829..21105  
 /standard\_name="170359"  
 repeat\_region complement(23518..23829)  
 /rpt\_family="AluSg"  
 repeat\_region 24089..24348  
 /rpt\_family="MIR"  
 repeat\_region 24297..24359  
 /rpt\_family="L2"  
 repeat\_region complement(25626..25925)  
 /rpt\_family="AluSg"  
 repeat\_region 25976..26107  
 /rpt\_family="MIR"  
 repeat\_region 26108..26395  
 /rpt\_family="AluSx"  
 repeat\_region 26396..26417  
 /rpt\_family="(TAA)n"

Query Match 10.9%; Score 218.4; DB 8; Length 181174;  
 Best Local Similarity 74.1%; Pred. No. 2.3e-35;  
 Matches 304; Conservative 0; Mismatches 101; Indels 5; Gaps 2;

QY 287 ATTCAACAAGTAGTACGCTGCTATTTGAAGATTTAAAGCTATTTATTAACAACATTTAAT 346  
 |||||  
 Db 47159 ATTGACAATGCAATTCGTATCTGAACAGCTCCTCATTTGAAGAAGCAAAAGATTA 47100  
 QY 347 AAAATTTAAACTATACACTTAATTTAAAGCTTTGAATGGGCGAGCGAG 406  
 |||||  
 Db 47099 ACAATATGATCAATTAATTAATGATTAATAAACTGATTAAGAGTAGGCGGACAG 47040  
 QY 407 TAGCTCGCTGCTGAATCCCAACACTTTGGGAGGCCAAGGTGGGCGGATCACTGAGTCT 466  
 |||||  
 Db 47039 TGCTCAGCGCTGTAATCCAGACCTTTGGAGGCGGAGGTGGGCGGATCACTGAGTCT 46980  
 QY 467 AGGAGTTTAAAGACGAGCTGGCCCAACATGTGAAACCTGTCTTAATAAAACGAAA 526  
 |||||

Db 46979 AGGAGTTGAGACCAAGCTTAAAGCAATGATGTAACCTGTCTCTAATAAATACAAA 46920  
Qy 527 ATTACCAAGTGTGGTGCATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGGA 586  
Db 46919 ATTACCGGCTGTGTGGTGCACCGCTGTAGTCCCACTACTCAGAGGTTGAGGAGGA 46860  
Qy 587 GAATGCTTGAACCTAGAGGTTGAGGTTGACAGTAAACCGAGA---TGTCACTGCACCTCC 643  
Db 46859 GTGTGCTGGAACCTGTGGAGGAGGAGGTTGACAGTAAACCGAGTTGTGTCACTGCACCTCC 46800  
Qy 644 AGCCT--GGCACAAGACAGACACTCATTAAGACAAACAAAGCTTTGAAA 691  
Db 46799 ACCTGGGCGACAGAGTAAGACTCATCTCAAAAAAAATGATATA 46750

RESULT 45  
AC161061/c 187825 bp DNA linear HTG 11-MAY-2005  
LOCUS Pan troglodytes chromosome UNKNOWN clone CH251-489A6, WORKING DRAFT  
DEFINITION  
SEQUENCE, 9 unordered pieces.  
AC161061  
AC161061.1 GI:63055279  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
KEYWORDS  
Pan troglodytes (chimpanzee)  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Pan.  
1 (bases 1 to 187825)  
Wilson, R.K.  
The sequence of Pan troglodytes clone  
Unpublished  
2 (bases 1 to 187825)  
Wilson, R.K.  
Direct Submission  
Submitted (06-MAY-2005) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 187825)  
Wilson, R.K.  
Direct Submission  
Submitted (11-MAY-2005) Genetics, Genome Sequencing Center, 4444  
Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@wustl.edu  
Project Information  
Center project name: C\_AB0489A06  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Sequencing vector: Plasmid; 100%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 181623 bases at least Q40  
Consensus quality: 182527 bases at least Q30  
Consensus quality: 183339 bases at least Q20  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1599: contig of 1599 bp in length  
\* 1600 1699: gap of unknown length  
\* 1700 3334: contig of 1535 bp in length  
\* 3235 3334: gap of unknown length

3335 4681: contig of 1347 bp in length  
4682 4781: gap of unknown length  
4782 7438: contig of 2657 bp in length  
7439 7538: gap of unknown length  
7539 21710: contig of 14172 bp in length  
21711 21810: gap of unknown length  
21811 38270: contig of 16460 bp in length  
38271 38370: gap of unknown length  
38371 73523: contig of 35153 bp in length  
73524 73623: gap of unknown length  
73624 115929: contig of 42306 bp in length  
115930 116029: gap of unknown length  
116030 187825: contig of 71796 bp in length.

FEATURES  
source  
1. 187825  
Location/Qualifiers  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/chromosome="UNKNOWN"  
/clone="CH251-489A6"  
1. 1599  
/note="assembly\_name:Contig550"  
1600. 1699  
/estimated\_length=unknown  
1700. 3234  
/note="assembly\_name:Contig572"  
3235. 3334  
/estimated\_length=unknown  
/note="assembly\_name:Contig584"  
4682. 4781  
/estimated\_length=unknown  
4782. 7438  
/note="assembly\_name:Contig587"  
7439. 7538  
/estimated\_length=unknown  
7539. 21710  
/note="assembly\_name:Contig588  
clone\_end:SP6  
vector\_side:right"  
21711. 21810  
/estimated\_length=unknown  
21811. 38370  
/note="assembly\_name:Contig589"  
38271. 38370  
/estimated\_length=unknown  
38371. 73523  
/note="assembly\_name:Contig590"  
73524. 73623  
/estimated\_length=unknown  
73624. 115929  
/note="assembly\_name:Contig591"  
115930. 116029  
/estimated\_length=unknown  
116030. 187825  
/note="assembly\_name:Contig592"

ORIGIN  
Query Match 10.9%; Score 218.4; DB 14; Length 187825;  
Best Local Similarity 59.1%; Pred. No. 2.3e-35;  
Matches 447; Conservative 0; Mismatches 301; Indels 8; Gaps 4;  
Qy 1 TGATTCACACGCTTGGCTTCCCAAGTGTGGATTACAGGCGGTGAGCCACGACGCTG 60  
Db 16226 TGATTCGCGCCCGCTCGCGCTCTCTTAAGTGTGGATTACAGGCGGTGAGCCACGCGCGCAG 16167  
Qy 61 GTTCGAATGCTTTTATTATTAAGAGACACATGGGCTTAATCTGTCTTATTGAC 120  
Db 16166 GCC--AGGCTGAATTAATTTTAAATGATTTTACACTGGAGAAAGATTGACCAAAATC 16109  
Qy 121 AGACTTGTATGAGTCAATCCCAATGCTGCACCTTACTGAACGCGCTTAATGACTTAC 180  
Db 16108 ATTCATTTGCAATTTTCACTGTAATGACATATAAATCTTTTACTTATATATTAAGCTTAA 16049

```
OY 181 TCTCTCAGCTGCTCTTCTGCAATGTAAGGTGGAATATGATGCTTTCAAGAGNA 240
DB 16048 AGTATAAAAGATTGTTATGATTTTATGACATATCTACTTTTCATTTGTGAGAACCC 15989
OY 241 TAAACCTATGAAAAGTGTGAGATAGTGTGATATGAAATAGAGATTCAACAAGTAG 300
DB 15988 TAGAGAAAAGGTGTACAAAAGAAAACCCCCAAAATCTCTACTTTAATGAAATTA 15929
OY 301 TAGCTGCTATTGAAGATTTAAGAGTTATTATTAACAATTTAATTAATAATTTTAAAAAC 360
DB 15928 AAGCTAATGATGTAGCTGCAATTCCTATTATTAGTTCTCTAAAATGCTAATTCCTTAT 15869
OY 361 TAATACACTTAATTATTTAAAGAGCTTGAATGGGCCAGGCCAGTAGCTCCTGCTGT 420
DB 15868 TAATTACTGTCTTATTATTAATAACTGT-CTGGCCGGGCATGTGTGACACCTGT 15811
OY 421 AATCCCAACACTTTGGGAGGCCAAGGTGGCGGATCACCCTGAAGTCAGAGTTTAAGACC 480
DB 15810 AATCCAGCACTTTGGAGAGGTGAGGTGGCGGATCACCCTGAAGTCGAGATTCAAGAGC 15751
OY 481 AGCTTGGCCAATGTGTAACCCCTGTCTCTACTAATAAAGCAAAAATTAGCCAGGTGG 540
DB 15750 AGCTTGGCTTAACATGTGTGAACCCCTGTCTACTAATAAAGCAAAAATTAGCTGGGCATG 15691
OY 541 GTGGCATGACCTGTAGTCCCACTACTCAGAGGTTGAGGGAGAGAAATGCTTGAACC 600
DB 15690 GTGTGGGCACTGTATATTCAGCTACTGGGAGGCTGAGGCAAGAAATGCTTGAACC 15631
OY 601 TAGAGGTGAGAGTTGCACTAACCCGAGAT--GTCACTGCACTCCAGCTT-GGCAACAG 656
DB 15630 CAGGAGGTGAGAGTTGCACTGAGGCAAGATGCGCACTGCACTCCAGCTGGGTAAACAG 15571
OY 657 AGCAAGACTTCATTAAGACAAACAAGCTTTGAATTTGTAAATGAGTTGTACTATCT 716
DB 15570 ACCGAGACTTCATCTAAAAAAGAAAAAACAACAAAAAATCTTTTGGGGGGGTATAT 15511
OY 717 TCATTTAAGAAATTCATCTTTGTCATTTATTTTA 752
DB 15510 GACCTTAATTTAGTGTCTAATATATATCTTAATATA 15475
```

Search completed: January 22, 2006, 13:37:24  
Job time : 6828.68 secs





93	204.8	10.2	32152	9	ADB32581	Adb32581 Human nov
94	204.8	10.2	32152	10	ADB93121	Adb93121 Human col
95	204.8	10.2	32204	4	AAK89019	Aak89019 Human dig
96	204.8	10.2	32204	4	AAK91533	Aak91533 Human dig
97	204.8	10.2	32204	4	AA157790	Aa157790 Human col
98	204.8	10.2	32204	5	AA839620	Aa839620 Genomic s
99	204.8	10.2	32204	6	AB89967	Ab89967 Genomic D
100	204.8	10.2	32204	9	ADB32580	Adb32580 Human nov

ALIGNMENTS

RESULT 1  
AAV83939 standard; DNA; 80595 BP.  
AAV83939;  
03-MAR-1999 (first entry)  
HC-contig derived from normal human chromosome 10q25.2 region.

Yeast artificial chromosome; YAC; probe: eukaryotic chromosome;  
neocentromere; replication; extra-chromosomal element; segregation;  
cell division; artificial chromosome; gene therapy; mardel(10);  
human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.  
Homo sapiens.  
WO9851790-A1.  
19-NOV-1998.  
13-MAY-1998; 98WO-AU000352.  
13-MAY-1997; 97AU-00006784.  
26-AUG-1997; 97AU-00008791.  
(AMRA-) AMRAD OPERATIONS PTY LTD.  
Choo K, Du Sart D, Cancelli MR;  
MPI; 1999-009773/01.  
New isolated nucleic acid comprising neocentromere sequences from  
eukaryotic chromosome - used to produce replicable, segregating  
artificial chromosomes that can carry large amounts of DNA for gene  
therapy.

Claim 8; Fig 6; 540bp; English.

The present sequence represents the HC-contig derived from normal human  
chromosome 10, 10q25.2 region. This region can be naturally mutated to  
produce an unusual chromosomal marker referred to as mardel(10). The  
mardel(10) marker is multicisally stable and contains a functional  
neocentromere at a location regarded as non-centromeric. This  
neocentromere maps to q25.2 on chromosome 10. The specification describes  
nucleic acid sequences derived from a eukaryotic chromosome, including a  
neocentromere or its functional derivative or hybrid, that are able, in a  
compatible cell, of replicating, acting as extra-chromosomal element and  
segregating during cell division. The sequences can be used to construct  
artificial chromosomes for use in gene therapy comprising a replicable,  
segregating nucleic acid that confers a specific phenotype on cells.  
Human artificial chromosomes can propagate in human cells and carry large  
amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,  
they are not mutagenic. The artificial chromosomes are also useful for  
generation of transgenic plants and animals, in production of proteins  
and to make diagnostic reagents, e.g. for expression of cytokines,  
receptors and growth factors, or to increase the copy number of a gene in  
a cell. The constructs may also be used for functional and structural  
analysis of chromosomes

Seq	Sequence	80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;
Query	Match	100.0%; Score 2001; DB 2; Length 80595;
Best	Local Similarity	100.0%; Pred. No. 0;
Matches	2001; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	TGATCCACACAGCTTGGCTCCCAAGTGCTGGAGTTACAGGCGTGAAGCCACACGCTTG 60
Db	39000	TGATCCACACAGCTTGGCTCCCAAGTGCTGGAGTTACAGGCGTGAAGCCACACGCTTG 39059
Qy	61	GTGCAATGCTTTTATTTATTTGAAGACAAATAGGCGCTTAAATCTGCTTTATTTGAC 120
Db	39060	GTGCAATGCTTTTATTTATTTGAAGACAAATAGGCGCTTAAATCTGCTTTATTTGAC 39119
Qy	121	AGACTTTATGAGGCAATCCCAATGCTGCACCTTACCTGAACGGCTTAAATGACTTAC 180
Db	39120	AGACTTTATGAGGCAATCCCAATGCTGCACCTTACCTGAACGGCTTAAATGACTTAC 39179
Qy	181	TCTCTGAGCTGCTTTCTGCAATGTAAGGTGAATTAATGATGCTTTCAAGAGGAA 240
Db	39180	TCTCTGAGCTGCTTTCTGCAATGTAAGGTGAATTAATGATGCTTTCAAGAGGAA 39239
Qy	241	TAAACCTTGAAGAGTGTGAGATAGGTTGATATGAATTAAGATTTCAACAGTAG 300
Db	39240	TAAACCTTGAAGAGTGTGAGATAGGTTGATATGAATTAAGATTTCAACAGTAG 39299
Qy	301	TAGCTGATTTGAATTTAAGATTTATTTATTTACACTAATTTAATTAATTTAAAAAC 360
Db	39300	TAGCTGATTTGAATTTAAGATTTATTTATTTATTTACACTAATTTAATTAATTTAAAAAC 39359
Qy	361	TAAATCACTTAATTTAATTAAGAGCTTTGAATTTGAATTTGAATTTGAATTTGAATTT 420
Db	39360	TAAATCACTTAATTTAATTAAGAGCTTTGAATTTGAATTTGAATTTGAATTTGAATTT 39419
Qy	421	TAATCCCAACCTTTGGAGGCGCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGAC 480
Db	39420	TAATCCCAACCTTTGGAGGCGCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGAC 39479
Qy	481	AGCTGGCCCAACATGGTGAACCCCTGCTCTACTAATAAAGCAAAATTTAGCAGGTGTG 540
Db	39480	AGCTGGCCCAACATGGTGAACCCCTGCTCTACTAATAAAGCAAAATTTAGCAGGTGTG 39539
Qy	541	GTGGCATGACCTGTGATCCCAACTACAGAGGTTGAGGAGGAAATTCCTGAACC 600
Db	39540	GTGGCATGACCTGTGATCCCAACTACAGAGGTTGAGGAGGAAATTCCTGAACC 39599
Qy	601	TAGAGGTGAGGTTGCACTTAACCCGAGATGTCACCTCAGGCTGGCAACAGAGA 660
Db	39600	TAGAGGTGAGGTTGCACTTAACCCGAGATGTCACCTCAGGCTGGCAACAGAGA 39659
Qy	661	AGACTCCATTAAGACAAACAAAGCTTTGAATTTGTAAATGATGTTGACTTATCTTCAT 720
Db	39660	AGACTCCATTAAGACAAACAAAGCTTTGAATTTGTAAATGATGTTGACTTATCTTCAT 39719
Qy	721	TTAAGAAATTCATCTTTGTTCAATTTATTTTTCATGACATGAGGCTTCCAGCAATTTT 780
Db	39720	TTAAGAAATTCATCTTTGTTCAATTTATTTTTCATGACATGAGGCTTCCAGCAATTTT 39779
Qy	781	AATTAAGCCTCAAGATTTATGTCATGCTGCTATGATTAACAAATTTATTTGCTTAA 840
Db	39780	AATTAAGCCTCAAGATTTATGTCATGCTGCTATGATTAACAAATTTATTTGCTTAA 39839
Qy	841	TAAATTTCTGCTTTTATTTAAGAAATGCTCCAGAAACGTTTGTACCAACAT 900
Db	39840	TAAATTTCTGCTTTTATTTAAGAAATGCTCCAGAAACGTTTGTACCAACAT 39899
Qy	901	ACACTGACTTTACAAATACATGATGATTTGCAACAGTTGCAAGATTTTCAAAAGAT 960
Db	39900	ACACTGACTTTACAAATACATGATGATTTGCAACAGTTGCAAGATTTTCAAAAGAT 39959
Qy	961	TTTCATTTGAGAGGGGCCCATTTGGGTATTTAGATTTTAAGATTTGAACATGCTTTGT 1020
Db	39960	TTTCATTTGAGAGGGGCCCATTTGGGTATTTAGATTTTAAGATTTGAACATGCTTTGT 40019

```
OY 1021 TCTGTTTTCTGGCTTCTGGAGAGAGACATGAATTCAGTTAGACCTTGATTT 1080
    |||||
DB 40020 TCTGTTTTCTGGCTTCTGGAGAGAGAGACATGAATTCAGTTAGACCTTGATTT 40079
OY 1081 TCTTTATCTCTTCAATTTCAATACAGAAAGTCTTCATATGACAGTGGTGCAGTCAAT 1140
    |||||
DB 40080 TCTTTATCTCTTCAATTTCAATACAGAAAGTCTTCATATGACAGTGGTGCAGTCAAT 40139
OY 1141 CAAAAAGAAAGAAACAGTTTCTTGTTTTTAATTTTCAACCGGAAAGAAAGGACCCA 1200
    |||||
DB 40140 CAAAAAGAAAGAAACAGTTTCTTGTTTTTAATTTTCAACCGGAAAGAAAGGACCCA 40199
OY 1201 TTTTGTCCGCTCAATTTAGCCAGTGCATGACTTAGAGAGAGAGAGAGAGAGAGG 1260
    |||||
DB 40200 TTTTGTCCGCTCAATTTAGCCAGTGCATGACTTAGAGAGAGAGAGAGAGAGAGG 40259
OY 1261 GTGTATACACAGGCTCTTCAATTAATCTCACAGAGAGAGAGAGAGAGAGAGAGG 1320
    |||||
DB 40260 GTGTATACACAGGCTCTTCAATTAATCTCACAGAGAGAGAGAGAGAGAGAGAGG 40319
OY 1321 TGAGAGAAATGGCTCGAGAGAGATGAACAATCTCACAGAGAGAGAGAGAGAGAGG 1380
    |||||
DB 40320 TGAGAGAAATGGCTCGAGAGAGATGAACAATCTCACAGAGAGAGAGAGAGAGAGG 40379
OY 1381 TGTATTAAGACCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 1440
    |||||
DB 40380 TGTATTAAGACCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40439
OY 1441 TCTACGTCAGAGAGAGAGAGAGAGAGATTAATTTAGTGAATTTAGAAATTTTAAATTA 1500
    |||||
DB 40440 TCTACGTCAGAGAGAGAGAGAGAGATTAATTTAGTGAATTTAGAAATTTTAAATTA 40499
OY 1501 TTCTAAAGTCATGACACAGCTTAATTAATAAGATAGTGGTGAAGGTGCTGAATTAAC 1560
    |||||
DB 40500 TTCTAAAGTCATGACACAGCTTAATTAATAAGATAGTGGTGAAGGTGCTGAATTAAC 40559
OY 1561 TGGATTTTACCAACCCCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 1620
    |||||
DB 40560 TGGATTTTACCAACCCCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40619
OY 1621 TCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 1680
    |||||
DB 40620 TCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40679
OY 1681 TTAACCAATACAAATGACTAGTGTCTGTCTCTTTATTAAGACACTAAATTTTGATCC 1740
    |||||
DB 40680 TTAACCAATACAAATGACTAGTGTCTGTCTCTTTATTAAGACACTAAATTTTGATCC 40739
OY 1741 AATTAATTAATCTGTCCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 1800
    |||||
DB 40740 AATTAATTAATCTGTCCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40799
OY 1801 TCAAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 1860
    |||||
DB 40800 TCAAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40859
OY 1861 CAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 1920
    |||||
DB 40860 CAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40919
OY 1921 GTTGCATTTGAGACTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 1980
    |||||
DB 40920 GTTGCATTTGAGACTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40979
OY 1981 TCTAATTCAGTGAACACTTCA 2001
    |||||
DB 40980 TCTAATTCAGTGAACACTTCA 41000
```

RESULT 2  
AAV83940  
ID AAV83940 strand; DNA; 80240 BP.

```
XX AAV83940;
AC
DT 03-MAR-1999 (first entry)
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
XX Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
XX neocentromere; replication; extra-chromosomal element; segregation;
XX cell division; artificial chromosome; gene therapy; mardel(10);
XX human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
OS Homo sapiens.
XX
XX WO951790-A1.
XX
XX 19-NOV-1998.
XX
XX 13-MAY-1998; 98WO-AU000352.
XX
XX 13-MAY-1997; 97AU-00006784.
XX
XX 26-AUG-1997; 97AU-00008791.
XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Choo K, Du Sart D, Cancellia MR;
XX
XX WPI; 1999-009773/01.
XX
XX New isolated nucleic acid comprising neocentromere sequences from
XX eukaryotic chromosome - used to produce replicable, segregating
XX artificial chromosomes that can carry large amounts of DNA for gene
XX therapy.
XX
XX Claim 9; Fig 16A; 540bp; English.
XX
XX The present sequence represents the NC-contig derived from a mutated
XX human chromosome 10, 10q25.2 region. The sequence contains an unusual
XX chromosomal marker referred to as mardel(10). The mardel(10) marker is
XX mitotically stable and contains a functional neocentromere at a location
XX regarded as non-centromeric. This neocentromere maps to q25.2 on
XX chromosome 10. The specification describes nucleic acid sequences derived
XX from a eukaryotic chromosome, including a neocentromere or its functional
XX derivative or hybrid, that are able, in a compatible cell, of
XX replicating, acting as extra-chromosomal element and segregating during
XX cell division. The sequences can be used to construct artificial
XX chromosomes for use in gene therapy comprising a replicable, segregating
XX nucleic acid that confers a specific phenotype on cells. Human artificial
XX chromosomes can propagate in human cells and carry large amounts of DNA
XX (e.g. therapeutic genes), and, being extra-chromosomal, they are not
XX mutagenic. The artificial chromosomes are also useful for generation of
XX transgenic plants and animals, in production of proteins and to make
XX diagnostic reagents, e.g. for expression of cytokines, receptors and
XX growth factors, or to increase the copy number of a gene in a cell. The
XX constructs may also be used for functional and structural analysis of
XX chromosomes
XX
XX Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;
SQ
XX
XX Query Match 97.9%; Score 1959.4; DB 2; Length 80240;
XX Best Local Similarity 99.4%; Pred. No. 0; Mismatches 11; Indels 2; Gaps 2;
XX Matches 1988; Conservative 0;
OY 1 TGAATCACCAGGCTTGGCTCCCAAGTCTGGAGATTACAGGCGTAGAGCAACGAGCTTG 60
DB 38738 TGATCCACAGGCTTGGCTCCCAAGTCTGGAGATTACAGGCGTAGAGCAACGAGCTTG 38797
OY 61 GTGCAATGCTTATTTATTTGAAGAGACAATGGGCTTAAATGCTTCTATTATTAAC 120
DB 38798 GTGCAATGCTTATTTATTTGAAGAGACAATGGGCTTAAATGCTTCTATTATTAAC 38857
OY 121 AGACTTGATGAGTCAATCCCAATGCTGCACTTACGAGAGGCTTAAATGACTTAG 180
```

Db 38858 AGACTTGTGAGTCAATCCCATGCTGCACCTTACGACGGCTTTAAATGACTAG 38917  
QY 181 TCTCTCAGCTGTCTTTCTGATATGTAAAGTGAATTAATGATGGCTTTCAAGAGGAA 240  
Db 38918 TCTCTCAGCTGTCTTTCTGATATGTAAAGTGAATTAATGATGGCTTTCAAGAGGAA 38976  
QY 241 TAAACCTTAAGAAAGTGTGAGATAGTGTGATATGAAATTAAGATTTCAACAAGTAG 300  
Db 38977 TAAACCTTAAGAAAGTGTGAGATAGTGTGATATGAAATTAAGATTTCAACAAGTAG 39035  
QY 301 TAGCTGCTAATGAGATTTAAGATTTAATTAATTAATTAATTAATTAATTAATTAATTA 360  
Db 39036 TAGCTGCTAATGAGATTTAAGATTTAATTAATTAATTAATTAATTAATTAATTAATTA 39095  
QY 361 TTAATACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420  
Db 39096 TTAATACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 39155  
QY 421 AATCCCAACACTTTGGAGGGCCAGGTGGCGGATCACTGAGGTGAGAGTTTAAGACC 480  
Db 39156 AATCCCAACACTTTGGAGGGCCAGGTGGCGGATCACTGAGGTGAGAGTTTAAGACC 39215  
QY 481 ACCCTGGCCAACTGGTGAACCTGTGTCTACTAATAAAACGCAAAATTAAGCCAGGTGTG 540  
Db 39216 ACCCTGGCCAACTGGTGAACCTGTGTCTACTAATAAAACGCAAAATTAAGCCAGGTGTG 39275  
QY 541 GTGGCATGCACTGTATGTCCCACTACTGAGAGGTGAGGGAGGAAATGGCTTGAACC 600  
Db 39276 GTGGCATGCACTGTATGTCCCACTACTGAGAGGTGAGGGAGGAAATGGCTTGAACC 39335  
QY 601 TAGAGAGTGAAGGTTCAGATTAACCCGAGATGTCACTGCACTCCAGCTGGCAACAGAGA 660  
Db 39336 TAGAGAGTGAAGGTTCAGATTAACCCGAGATGTCACTGCACTCCAGCTGGCAACAGAGA 39395  
QY 661 AGACTCCATTAAGACAAACAAAGCTTTGAATTTGTAAATGATGTATCTATCTTCAAT 720  
Db 39396 AGACTCCATTAAGACAAACAAAGCTTTGAATTTGTAAATGATGTATCTATCTTCAAT 39455  
QY 721 TTAAGAAATTCATCTTTGTTCATTTTATTTTACTTGAACATGAGCTTCCAGAAATTTT 780  
Db 39456 TTAAGAAATTCATCTTTGTTCATTTTATTTTACTTGAACATGAGCTTCCAGAAATTTT 39515  
QY 781 AATTAAGCCCTCAGATTTTATGTCACTGGCTATGATTAACAAATTTTGTCTAAAA 840  
Db 39516 AATTAAGCCCTCAGATTTTATGTCACTGGCTATGATTAACAAATTTTGTCTAAAA 39575  
QY 841 TTAATATCTTCTCTCTTTTAAAGAAATGTCTCCCTAGAAACGGTTTGTACCAAAAT 900  
Db 39576 TTAATATCTTCTCTCTTTTAAAGAAATGTCTCCCTAGAAACGGTTTGTACCAAAAT 39635  
QY 901 ACACTGACTTTACCAAAATCAGATCTGATTTGGCAACAGTTGCAGATTTTCAAAAAT 960  
Db 39636 ACACTGACTTTACCAAAATCAGATCTGATTTGGCAACAGTTGCAGATTTTCAAAAAT 39695  
QY 961 TTTCAATTTGAGAGGGGCCCATTTGGTTATTTAGATTTCTAAGAACTGAAATGCTTTGT 1020  
Db 39696 TTTCAATTTGAGAGGGGCCCATTTGGTTATTTAGATTTCTAAGAACTGAAATGCTTTGT 39755  
QY 1021 TCTGTTTTCTGGCTTCTGGAGAGAGAGACATGAATTCAGTTAGCACCTTGATTTT 1080  
Db 39756 TCTGTTTTCTGGCTTCTGGAGAGAGAGACATGAATTCAGTTAGCACCTTGATTTT 39815  
QY 1081 TCTTTATCTTCAATTAATACAGAAATGCTTCAATATGACAGTGTGTCACTGTCAT 1140  
Db 39816 TCTTTATCTTCAATTAATACAGAAATGCTTCAATATGACAGTGTGTCACTGTCAT 39875  
QY 1141 CAAAAGAAAGAGAAAGCTTTCTGGTTTAAATTTTCAACCGGAAAGAAAGCACCA 1200  
Db 39876 CAAAAGAAAGAGAAAGCTTTCTGGTTTAAATTTTCAACCGGAAAGAAAGCACCA 39935  
QY 1201 TTTTGTTCGCGCTAATTTAGCAGTGCATGACTTGAAGAGCAGACAGTCTTTGAAGGC 1260  
Db 39936 TTTTGTTCGCGCTAATTTAGCAGTGCATGACTTGAAGAGCAGACAGTCTTTGAAGGC 39995

QY 1261 GTGGTAACACAGGTCTTCAATTAATCTCCAGCAGAGCTTGCACTTCAATGCTTAGGC 1320  
Db 39996 GTGGTAACACAGGTCTTCAATTAATCTCCAGCAGAGCTTGCACTTCAATGCTTAGGC 40055  
QY 1321 TGAAGAAATATGCTCAGAGAAATGAACAATCTCCAGAGCCCTTAATCTAAGAGCCAGG 1380  
Db 40056 TGAAGAAATATGCTCAGAGAAATGAACAATCTCCAGAGCCCTTAATCTAAGAGCCAGG 40115  
QY 1381 TGTATTAAGCAAGTCAAGAGGGGTGAGAACTAAGCTTCTGTAATCTCCACTTCTT 1440  
Db 40116 TGTATTAAGCAAGTCAAGAGGGGTGAGAACTAAGCTTCTGTAATCTCCACTTCTT 40175  
QY 1441 TCTACGTCAAGAGCCCAAGCTGATTTATTTAGTTGGAATTTAGAAATTTTAAATA 1500  
Db 40176 TCTACGTCAAGAGCCCAAGCTGATTTATTTAGTTGGAATTTAGAAATTTTAAATA 40235  
QY 1501 TTTCTAAGTCAATGAACAGCCCTTAATTAAGAAATGCTGTGTAAGGTGCGAAATAC 1560  
Db 40236 TTTCTAAGTCAATGAACAGCCCTTAATTAAGAAATGCTGTGTAAGGTGCGAAATAC 40295  
QY 1561 TCGATTTTACCAACCCCTCTTCTGAGAGAACCATTAATGAAATCTGTACATGTTTAC 1620  
Db 40296 TCGATTTTACCAACCCCTCTTCTGAGAGAACCATTAATGAAATCTGTACATGTTTAC 40355  
QY 1621 TCTACCAACGAACCTGTTTCTTAATGAGAGAAACAGAGCCCAAGTATTAACATAC 1680  
Db 40356 TCTACCAACGAACCTGTTTCTTAATGAGAGAAACAGAGCCCAAGTATTAACATAC 40415  
QY 1681 TTAACCAATACAAATGACTATGCTGTGCTCTTTTAAAGCACTAAATTTGATCC 1740  
Db 40416 TTAACCAATACAAATGACTATGCTGTGCTCTTTTAAAGCACTAAATTTGATCC 40475  
QY 1741 AATTAATTAATCTGTCCATTAAGAGAGTTTCCCTAATGATCTGTTCACTTGTCCCT 1800  
Db 40476 AATTAATTAATCTGTCCATTAAGAGAGTTTCCCTAATGATCTGTTCACTTGTCCCT 40535  
QY 1801 TCAAGGGGCAAGTGTCCCTTACACATAGCTAAATGGAATTTCTTTCAATCAATACC 1860  
Db 40536 TCAAGGGGCAAGTGTCCCTTACACATAGCTAAATGGAATTTCTTTCAATCAATACC 40595  
QY 1861 CAGAGGGCAAACTTAATATGCTGTGAATGAACATTTGCTGTTCACATCTCAGAGCACT 1920  
Db 40596 CAGAGGGCAAACTTAATATGCTGTGAATGAACATTTGCTGTTCACATCTCAGAGCACT 40655  
QY 1921 GTTGCAATTTGAGCTTCTGAGGGGCCACCCAGAGCTAATATGCTGAGATGTTAACTCA 1980  
Db 40656 GTTGCAATTTGAGCTTCTGAGGGGCCACCCAGAGCTAATATGCTGAGATGTTAACTCA 40715  
QY 1981 TCTAATTCAGTGAACATTTCA 2001  
Db 40716 TCTAATTCAGTGAACATTTCA 40736

RESULT 3  
AAC6548  
ID AAC6548 standard; DNA; 121162 BP.  
XX  
AC AAC6548;  
XX  
DT 19-FEB-2001 (first entry)  
XX  
DE Human kinesin-like protein HKLP; KIF1; cell division; cancer;  
XX Human; kinesin-like protein; HKLP; KIF1; cell division; cancer;  
XX intracellular transport; neurological disorder; infertility;  
XX KM biallelic marker; spontaneous abortion; neonatal chromosome disorder;  
XX aneuploidy; ds.  
XX  
XX Homo sapiens.  
XX  
XX PN W0200063375-A1.  
XX

PD 26-OCT-2000.  
XX  
PF 20-APR-2000; 2000MO-IB000562.  
XX  
PR 20-APR-1999; 99US-0130217P.  
XX  
PA (GEST ) GENSET.  
XX  
PI Bougueleret L, Dufaire-Gare I, Grel P;  
XX WPI; 2000-665242/64.  
DR  
XX  
XX An isolated or purified human kinesin-like protein (HKLP) encoding  
PT polynucleotide used to detect HKLP polynucleotides in a sample comprises  
PT a contiguous span of at least 12 nucleotides.  
XX  
XX Claim 1; Page 143-175; 1999P; English.  
XX  
XX The present invention describes the coding and protein sequences of the  
CC human kinesin-like protein HKLP. It is thought that the protein could be  
CC involved in neurological disorders, infertility, spontaneous abortion,  
CC neonatal chromosome disorders, aneuploidy and cancers. This is due to its  
CC function in the movement of microtubules. The protein shows homology to  
CC the murine KIF1A and KIF1B proteins. The sequences disclosed in the  
CC invention can be used in the isolation of similar human proteins and in  
CC vector production. In addition, the allelic markers shown can be used  
CC in disease diagnosis and population studies  
XX  
SQ Sequence 121162 BP; 33272 A; 24108 C; 25842 G; 37919 T; 0 U; 21 Other;  
Query Match 11.1%; Score 221.2; DB 3; Length 121162;  
Best Local Similarity 83.8%; Pred. No. 2.2e-36;  
Matches 263; Conservative 0; Mismatches 48; Indels 3; Gaps 1;  
QY 388 TGAATGGGCGCAGGAGTCTCTGCTGTATCCCAACATTGGAGGCCAAGT 447  
DB 25335 TGAATGGGCGCAGGAGTCTCTGCTGTATCCCAACATTGGAGGCCAAGT 25394  
QY 448 GGGCGGATCACTGAGGTCAAGAGTTTAAAGCAGAGCTGGCCAACTGGTGAACCTGT 507  
DB 25395 GGGCGGATCACTGAGGTCAAGAGTTTAAAGCAGAGCTGGCCAACTGGTGAACCTGT 25454  
QY 508 CTCTATAAAAAAGCAAAATTTAGCAGGTGTGTGAGTCACTGCTGTATCCCAACATTGGAGGCCAAGT 567  
DB 25455 CTCTATAAAAAAGCAAAATTTAGCAGGTGTGTGAGTCACTGCTGTATCCCAACATTGGAGGCCAAGT 25514  
QY 568 TCAGAGGTTGAGGAGGAGAAATTTGTTGAACCTAGAGGTGAGGTTGCAGTAACTCA 627  
DB 25515 TCAGAGGTTGAGGAGGAGAAATTTGTTGAACCTAGAGGTGAGGTTGCAGTAACTCA 25574  
QY 628 GAT--GTCACTGCACTCCAGCTGGCAAGCAAGACTCATTAAGCAACAAAGC 684  
DB 25575 GATGGGCGCACTGCACTCCAGCTGGCAAGCAAGACTCATTAAGCAACAAAGC 25634  
QY 685 TTTGAATTTGTGA 698  
DB 25635 AAATTAATTTCTGA 25648  
RESULT 4  
ADQ17382  
ID ADQ17382 standard; DNA; 153170 BP.  
AC ADQ17382;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 199.  
XX  
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KW ds.  
XX  
OS Homo sapiens.

XX  
PN MO2004048938-A2.  
XX  
XX 10-JUN-2004.  
PD  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX  
XX Aziz N, Ginsburg WM, Zlotnik A;  
PI WPI; 2004-441208/41.  
DR  
XX  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 199; 210pp; English.  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 153170 BP; 39941 A; 37835 C; 36509 G; 38885 T; 0 U; 0 Other;  
Query Match 10.8%; Score 215.6; DB 12; Length 153170;  
Best Local Similarity 77.6%; Pred. No. 3.4e-35;  
Matches 287; Conservative 0; Mismatches 79; Indels 4; Gaps 2;  
QY 311 TGAAGTTTAAAGGTTTAAATTTATTAACA-CTATTATAAAATTTAAAGTAATCACT 369  
DB 73952 TTAATAATTTAAATTTATCTCATTTATATATGCAATTTTCTAGATTACAGAGACCTCACT 74011  
QY 370 TAAATTATTAAAGAGCTTTGAATGGCCAGGCGCAGTAGCTCTGCTGTATCCCAAC 429  
DB 74012 TTTCCGCTTATTAATTAATTCAGTAAAGCCGGGCGCAGGTGCTCATGCTGTATCCCAAC 74071  
QY 430 ACTTTGGAGGCCAAGGTGGGCGGATCACCTGAGGTGAGAGTTTAAAGCAAGCTGGCC 489  
DB 74072 ACTTTGGAGGCCAAGGTGGGCGGATCACCTGAGGTGAGAGTTTAAAGCAAGCTGGCC 74131  
QY 490 AACATGTTGAACCCGCTCTCTAATAAAAGCAAAATTTAGCCAGGTGTGGTGCATGC 549  
DB 74132 AACATGTTGAACCCGCTCTCTAATAAAAGCAAAATTTAGCCAGGTGTGGTGCATGC 74191  
QY 550 ACCGTGAGTCCCAACTACTCAGAGGTTGAGGAGGAGATTGCTTGAACCTAGAGAGT 609  
DB 74192 ACCGTGAGTCCCAACTACTCAGAGGTTGAGGAGGAGATTGCTTGAACCTAGAGAGT 74251  
QY 610 GAGGTTGCAAGTAAACCCGAGAT--GTCACTGCACTTCAGCTGGCAACGAGCAAGACTC 666  
DB 74252 GAGGTTGCAAGTAAACCCGAGAT--GTCACTGCACTTCAGCTGGCAACGAGCAAGACTC 74311  
QY 667 CATTAAGACA 676  
DB 74312 CATTAAGACA 74321

RESULT 5  
AAS43104/c

II AAS43104 standard; DNA; 325791 BP.  
AC AAS43104;  
XX  
XX  
XX 18-DEC-2001 (first entry)  
XX  
XX Human Oestrogen receptor beta gene.  
XX  
XX Human: Oestrogen receptor beta; ERbeta; ds; SNP; chromosome 6q.25.1;  
KM single nucleotide polymorphism; cardiovascular disease;  
KM autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;  
KM osteoarthritis; osteoporosis; breast cancer; endometrial cancer.  
XX  
XX Homo sapiens.  
XX  
XX WO200162793-A2.  
XX  
XX 30-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US005360.  
XX  
XX 22-FEB-2000; 2000US-0183755P.  
XX 24-JAN-2001; 2001US-00768185.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;  
PI WPI, 2001-582041/65.  
XX P-PSDB; AAU27322.  
XX  
XX Estrogen receptor gene and protein polymorphisms useful for diagnosis of  
PT individuals at risk of developing bone disorders.  
XX  
XX Example 2; Fig 1; 245pp; English.  
XX  
XX The invention relates to a novel isolated peptide comprising or  
XX consisting of an amino acid sequence selected from an amino acid sequence  
XX of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of  
XX 10 amino acids), antibodies against them, nucleic acids encoding them  
XX (including vectors for transforming cells). The gene for human ERbeta is  
XX located on chromosome 6q.25.1. The variants are encoded by single  
XX nucleotide polymorphisms (SNP). The variant peptides and proteins can be  
XX used in assays to determine the biological activity of the protein, to  
XX raise antibodies, as a reagent in assays designed to quantitatively  
XX determine levels of the protein in biological fluids, to identify  
XX compounds that modulate receptor activity and to screen compounds for the  
XX ability to stimulate or inhibit interaction between the receptor protein  
XX and a target molecule that normally interacts with the receptor protein  
XX e.g. oestrogen. The antibody can be used to isolate the protein, to  
XX assess expression in disease states e.g. cardiovascular disease and  
XX autoimmune disease (e.g. systemic lupus erythematosus, arthritis,  
XX rheumatism and osteoarthritis), osteoporosis, breast cancer and  
XX endometrial cancer. In addition the antibodies can be used in  
XX pharmacogenomic analysis and inhibiting protein function, e.g. blocking  
XX the binding of the oestrogen receptor protein to a binding partner such  
XX as a ligand. The nucleic acids encoding the proteins can be used as  
XX probes, primers, chemical intermediates and in biological assays. The  
XX present sequence is the human ERbeta gene  
XX  
XX Sequence 325791 BP; 94098 A; 68292 C; 67970 G; 95431 T; 0 U; 0 Other;  
SQ  
Query Match 10.7%; Score 214.4; DB 4; Length 325791;  
Best Local Similarity 73.0%; Pred. No. 6.7e-35;  
Matches 305; Conservative 0; Mismatches 106; Indels 7; Gaps 2;  
QY 268 TGTTCATATGAAATGAAGATTCACAGTAGCTGCTATGAAGATTTAAGACTTA 327  
DB 257049 TGATTCGAAGATTCATATTTATCAACAAGCTTCAACAATTTAAATTAAGAG 256990  
QY 328 TTTATTTACAACCTATTTAATTAATTTAATTAATTAATTAATTAATTAATTAAGAGCTT 387  
DB 256989 ATACATCTCTATGAGAGCATTTAATAGCAAGAGTGACTTGAAGAAAGGCTTT 256930

QY 388 TGAAATGGCCAGGCGCAGTAGCTCTGCTGTAAATCCCAACATTTTGGAGCCAGGT 447  
DB 256929 ----CTGGCCATGCGCAGTAGCTTATGCTTGTATCCAGCAGCTTTGGAGGCTGAGGC 256874  
QY 448 GGGCGGATTCACCTGAGGTTCAGAGATTTAAAGACCACTGGCCCAACATGGTGAACCCCTGT 507  
DB 256873 GGGTGAATCCACTGAGGTTCAGAGATTTAAAGACCACTGGCCCAACATGGTGAACCCCTGT 256814  
QY 508 CTCTACTAATAAAGCAAAATTTAGCCAGTGTGTGTGAGCATGACCTGTAGTCCCAACTAC 567  
DB 256813 CTCTACCAAAATTAACAAATTTAGCTGGGTATGTGTGAGCAGCAGCTGTATCCAGCTAC 256754  
QY 568 TCAGAGGTTTGAAGGAGAGAAATTTGCTTGAACCTTAGAGGTGAGGTTGCAAGTAAACCGCA 627  
DB 256753 CTGGAGGCTGAGGAGGAGAAATTTGCTTGAACCTTAGAGGTGAGGTTGCAAGTAAACCGCA 256694  
QY 628 GAT--GTCACTGCACTCCAGCTTGGCAACAGAGCAAGCTCATTAAGAACAAACAAA 682  
DB 256693 GATCAAGCCCTGCACTCCAGCTTGGCAACAGAGCAAGCTCATTAAGAACAAACAAA 256636

## RESULT 6

ACCS7389 standard; cDNA; 2013 BP.

ID ACCS7389  
AC ACCS7389;  
XX  
XX 27-JUN-2003 (first entry)  
XX  
XX Human macroprotein 1503-8.8 encoding cDNA.  
XX  
XX Human; macroprotein; 1503-8.8; dementia; diabetes; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 911..1153  
XX FT /\*tag= a  
XX FT /product= "1503-8.8"  
XX  
XX CN1359929-A.  
XX  
XX 24-JUL-2002.  
XX  
XX 20-DEC-2000; 2000CN-00135135.  
XX  
XX 20-DEC-2000; 2000CN-00135135.  
XX  
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX  
XX Mao Y, Xie Y;  
XX  
XX WPI, 2002-733617/80.  
XX P-PSDB; ABP60188.  
XX  
XX Polypeptide-human macroprotein 1503-8.8 and polynucleotide for coding it.  
XX  
XX Claim 6; p25-26 (disclosure); 33pp; Chinese.  
XX  
XX The invention relates to a novel human macroprotein designated 1503-8.8.  
XX CC Also disclosed are the polynucleotide encoding the polypeptide, and the  
XX CC process for preparing the polypeptide using DNA recombination. The  
XX CC application of the polypeptide is in treating diseases such as dementia  
XX CC and diabetes. The current sequence represents the human macroprotein 1503  
XX CC -8.8 encoding cDNA  
XX  
XX Sequence 2013 BP; 622 A; 373 C; 391 G; 627 T; 0 U; 0 Other;  
SQ

Query Match 10.7%; Score 213.6; DB 6; Length 2013;  
Best Local Similarity 85.6%; Pred. No. 4.8e-35;  
Matches 250; Conservative 0; Mismatches 39; Indels 3; Gaps 1;





CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a bloodip;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182566A1, for which no sequence data was published

XX Sequence 59725 BP; 14028 A; 15992 C; 16067 G; 13499 T; 0 U; 139 Other;

SQ Query Match 10.6%; Score 212.2; DB 11; Length 59725;

Best Local Similarity 81.0%; Pred. No. 1.5e-34;  
Matches 260; Conservative 0; Mismatches 58; Indels 3; Gaps 1;

```
QY 364 TACACTTAATTAATTAAGAGCTTTGAAATGGCCAGCCAGTACTCTGCTTAAT 423
    |||||
DB 31622 TACACAAAATGTGTGAAAATGTGAGACGGCCGGGTGCTCACTGCTTAAT 31681

QY 424 CCCAACACTTTGGAGGCCAAGGTGGCGGATGATCCTGAGGTGAGGATTAGACGAGC 483
    |||||
DB 31682 CCTAGCACTTTGGAGGCCAAGCGGGTGGATCACTGAGGTGAGGATTAGACCATC 31741

QY 484 CTGGCCAACTAGGTGAAACCTGTCTTACTTAAACGCAAAATTAGCCAGGTGTGTG 543
    |||||
DB 31742 CTGGCCAACTAGGTGAAACCTGTCTTACTTAAACGCAAAATTAGCCAGGTGTGTG 31801

QY 544 GCATGCACTTGTATGCCAATCTACTAGAGGTTGAGGAGAGAAATGCTTAACTAG 603
    |||||
DB 31802 GCATGCACTTGTATGCCAATCTACTAGAGGTTGAGGAGAGAAATGCTTAACTAG 31861

QY 604 GAGGTGAGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
    |||||
DB 31862 GAGGTGAGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 31921

QY 661 AGACTCCATTAAGACAAACA 681
    |||||
DB 31922 AGACTCCATCTCAAAAATTA 31942
```

RESULT 9  
ADC87336/c  
ID ADC87336 standard; DNA; 108316 BP.

```
XX AC ADC87336;
XX DT 01-JAN-2004 (first entry)
XX DE Human GPCR gene SEQ ID NO:1789.
XX KW db; gene; human; GPCR;
XX KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX OS Homo sapiens.
XX PN EP1270724-A2.
XX PD 02-JUN-2003.
XX PF 18-JUN-2002; 2002EP-00013517.
XX PR 18-JUN-2001; 2001JP-00246789.
XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX DR WPI; 2003-315783/31.
XX DR P-PSDB; ADC87337.
```

PT New polynucleotide, useful for preparing a composition for treating a  
PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.

XX Claim 1; SEQ ID NO 1789; 28pp; English.

CC The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient  
CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The  
CC polynucleotide sequences shown in ADC8548-ADC87616 encode GPCR's of the  
CC invention.

XX SQ Sequence 108316 BP; 27970 A; 24159 C; 25607 G; 30580 T; 0 U; 0 Other;

Query Match 10.6%; Score 211.2; DB 10; Length 108316;

Best Local Similarity 67.9%; Pred. No. 2.7e-34;  
Matches 342; Conservative 0; Mismatches 153; Indels 9; Gaps 3;

```
QY 283 AAGATTTTCAACAGTAGTAGCTGTATTTGAAGATTTAAGATTATTTATCAACTATT 342
    |||||
DB 42571 AAAATTTTATCATGTATGATGATGATGATGATGATGATGATGATGATGATGAT 42512

QY 343 TATATAATTTTAAAACTAATACACTTAATTTTAAAGCTTTGAAATGGCCAGGC 402
    |||||
DB 42511 TTTTATCATCTTTTTCATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 42452

QY 403 GCATGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 462
    |||||
DB 42451 GAGGTGCTTCAAGCCCTGTATTTCCAGCACTTTGGAGGCCAAGGTGGCGGATCACCTTA 42392

QY 463 GGTCAAGAGTTTAAGACGAGCCCTGGCCACATGTGTAAACCTGTCTTACTTAAACCC 522
    |||||
DB 42391 GGTCAAGAGTTTGAAGCCAGCCCTGGCCACATGTGTAAACCTGTCTTACTTAAACCC 42332

QY 523 AAAATTTAGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 582
    |||||
DB 42331 AAAATTTAGCTGGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 42272

QY 583 AGGAGATTTGCTTGAACCTTGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 639
    |||||
DB 42271 AGGAGATTTGCTTGAACCTTGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 42212

QY 640 CTCCAGCCT-GGCAAGAGCAAGACTCCAT-----AAGACAAACAAAGCTTTGAATT 693
    |||||
DB 42211 CTCCAGCCTGGGCAAGAGCAAGACTCCATCTCAAAAATTTGATTAAGTCCAAAATTA 42152

QY 694 GTGTAAATGATGTGATCTTCACTTTTAAGAAATTCATCTTGTTCATTTATTTTTC 753
    |||||
DB 42151 TTTTCTAGAAAGATTTCATCAATCAATTAATTTCTATGTAGATACAGATTATTAATCT 42092

QY 754 TTGACATGAGAGCTTCCAGCAATT 777
    |||||
DB 42091 TGGCTTTGAAATTTTCAAGAAATT 42068
```

RESULT 10  
AAS41753/c  
ID AAS41753 standard; DNA; 11950 BP.

```
XX AC AAS41753;
XX DT 17-DEC-2001 (first entry)
XX DE Genomic sequence #69 encoding novel human enzyme polypeptide.
XX KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX KW ligase; hyperproliferative disorder; immunodeficiency disorder;
XX KW autoimmune disorder; neurological disorder; metabolic disorder;
XX KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
XX KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
```

KW anti arthritic; nephrotropic; anticoagulant; ds.  
XX Homo sapiens.  
XX MO200155301-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001MO-US001239.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-02311968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236527P.  
PR 29-SEP-2000; 2000US-0236527P.  
PR 29-SEP-2000; 2000US-0236567P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0255678P.



```
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229559P.
PR 05-SEP-2000; 2000US-0229551P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232357P.
PR 14-SEP-2000; 2000US-0232358P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WFI; 2001-476161/51.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
XX condition.
XX
XX Disclosure; SEQ ID NO 899; 859pp + Sequence Listing; English.
XX
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a genomic DNA of the invention
XX
XX
XX Sequence 11950 BP; 3159 A; 2617 C; 2790 G; 3384 T; 0 U; 0 Other;
SQ
Query Match 10.5%; Score 210.8; DB 4; Length 11950;
Best Local Similarity 59.6%; Pred. No. 2,4e-34;
Matches 409; Conservative 0; Mismatches 272; Indels 5; Gaps 3;
QY 1 TGAATCCAGGCTGGCTCCCAAGTGTGGATTACAGGCGTGGACCAAGCCCTG 60
DB 9406 TATCCACCCAGCTCAGCTCCCAAGTGTGGATTACAGGCGTGGACCAAGCCCTG 9347
QY 61 GTGGAATGCTTATTATTGAAGACAAATGGGCTTAAATCTGCTTATTGAC 120
DB 9346 GCAAAATTTTAAATATAAAAAACAC-CTAGAATATAAAAAATCAATCTTCA 9288
QY 121 AGACTTGAATGAGTCAATCCCAATGCTGCCACTTACGAAAGCGCTTAATGACTTG 180
DB 9287 ACTGTGCTGGCTCTCCCTTAAATCTCCGAAATGACCACTAGCCGAGTGGCTGTT 9228
QY 181 TCTCTCAGCTGTCTTTCGATATGTAGGTGAATATGATGCTTCAAGAGGAA 240
DB 9227 ATCCCTCAGAGCTGCTCTATGTATAAAAAATCAATATGTTTGAATTCAGTTTTC 9168
QY 241 TAACTTATGAAGAAGTGTAGGATAGTGTGATGTGAATTAAGGATTTCACAAATG 300
DB 9167 CTTACAACTTATATCACTAATATTTTATTATTAATATTAATTAATTAATGAATA 9108
QY 301 TAGCTGCTATTGAAGTTTAAGAGTTATTATTACAACTATTTAATTAATTAATAAC 360
```

DB 9107 TTTAATGACTGATATATTATTATTAATTTAAGGCGACATGTTAAACACTCAGACTCTA 9048  
QY 361 TTAATACCTTAATAATTATTAAAGAGCTTTGAATATGGCCCGGCGCATACTCTCTGCTGT 420  
DB 9047 GAGTACCAACTGTCCATATGAAATATGATGGAAAGACAGGCGCATGTGCTCAGCCTTC 8988  
QY 421 AATCCCAACACTTTGGAGAGCCCAAGGTGGGCGGATACCTGAGGTGAGGAGTTTAGACC 480  
DB 8987 AATCCAGCACTTTGGAGAGCTTGAAGTGGCGGATATCTTAGAGTTCAAGAGTTCAAGACC 8928  
QY 481 AGCTGCGCCAAACATGCTGAAACCTGTCTTACTAAAAACGAAAAATTAGCCAGGTGTG 540  
DB 8927 AGCTGCTCAACATGACCAAAATCCATCTCTAATAGAAATACAAAATAGCTGGGCGCTC 8868  
QY 541 GTGGCATGACCTGTGTGTCTCCCACTACTACGAGGAGTTGAGGAGGAGAAATGCTTTAACC 600  
DB 8867 GTGGCATGACCTGTGTGTCTCCCACTACTCGGAGGCTGAGGAGGAGATCCCTTTAACC 8808  
QY 601 TAGAGGTGAGGAGTTGAGTAAACCCGAG--ATGTCACCTCACTCCAGCCT-GGCAACAG 656  
DB 8807 CAGAGGTGAGGAGTTGAGTAAACCCGAGATCATTCCTGCACTCCAGCCTGGGCAACAG 8748  
QY 657 AGCAGACTCCATTAAGACAAACAAA 682  
DB 8747 AGTGAAGCTATGTCTCCAAAAAAA 8722

RESULT 12  
AAS29743/c  
ID AAS29743 standard; DNA; 11950 BP.  
XX  
XX AAS29743;  
DT 21-NOV-2001 (first entry)  
XX  
DE Human endocrine polypeptide encoding genomic DNA SEQ ID No 469.  
XX  
XX Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat;  
KM dog; chicken; sheep; immunosuppressive; antiarthritic; vasotrophic;  
KM antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KM cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KM ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;  
KM hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KM cerebrovascular disorder; nervous system disorder; bacterial infection;  
KM fungal infection; viral infection; ocular disorder; endocrine disorder;  
KM gastrointestinal disorder; renal disorder; respiratory disorder;  
KM wound healing; skin aging; organ transplantation; food preservative;  
KM tissue regeneration; anti-infertility.  
XX  
OS Homo sapiens.  
XX  
XX WO200155364-A2.  
PN  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001308.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214986P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225472P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234232P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234979P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 26-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241212P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.

08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-451936/48.  
XX  
PT Isolated polypeptide for treating, preventing and/or prognosing  
XX disorders of the endocrine system such as reproductive disorders,  
XX endocrine cancers and also for testing and detection e.g. diagnosis.  
PS  
XX Claim 1; SEQ ID NO 469; 604pp; English.

CC ocular disorders such as corneal infection, endocrine disorders such as  
CC premature labour and infertility, gastrointestinal disorders such as  
CC Crohn's disease, renal disorders such as glomerulonephritis and  
CC respiratory disorders such as asthma. The polypeptides can also be used  
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain  
CC organs before transplantation, to regenerate tissues and in chemotaxis.  
CC The polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Query Match 10.5%; Score 210.8; DB 5; Length 11950;  
Best Local Similarity 59.6%; Pred. No. 2.4e-34;  
Matches 409; Conservative 0; Mismatches 272; Indels 5; Gaps 3;

QY 1 TGATCCAGCAGCCTTGCCCTCCCAAGTCTGGATTACAGGCGTACCCAGCCTTG 60  
DB 9406 TGATCCAGCAGCCTTGCCCTCCCAAGTCTGGATTACAGTACGATGACGACCTTG 9347  
QY 61 GTGCATGCTTTATTTATTTGAAGACAACTAGGCTTTAATCTGCTTTATTTGAC 120  
DB 9346 GCATAATTTTAAATATAAAAAACAC-CTAGAAATAAAAGATAAAATCACTCTTAC 9288  
QY 121 AGACTTGATGAGATCAAAATCCCAATGCTGCACCTTACTGAAAGCGCTTAATGACTTAG 180  
DB 9287 ACTTGCTTGCTGCTCTCTCCCTTAAATCTCCCTGGAATACCACTACCTCCAGTGGCTGT 9228  
QY 181 TCTTCTCAGCTGCTCTTCTGCAATATGATGATGATGATGATGATGATGATGATGATG 240  
DB 9227 ATCCCTCCAGACGCTTCTGATGATATAAAAAATCAATATGATGATGATGATGATGAT 9168  
QY 241 TAACTCTATGAAAAGTGTGAGATATGCTGTTGATATGAAAATGAGATTTCAACAATG 300  
DB 9167 CTTACAACATTTATATCACTAATCACTTTTATTTATTTATTTATTTATTTATTTAT 9108  
QY 301 TAGCTGCTATGAAATTTAAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 360  
DB 9107 TTTATGACTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 9048  
QY 361 TAACTACTTAAATTTATTTAAGGCTTTGAAATGAGGCGAGGCGAGTACCTCTGCTGT 420  
DB 9047 GAGTGCACTGCTGCAATGAAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8988  
QY 421 AATCCCAACATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480  
DB 8987 AATCCCAACATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8928  
QY 481 AGCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
DB 8927 AGCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8868  
QY 541 GTGAGCATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGGAGGAGGAGGAGGAGGAGG 600  
DB 8867 GTGAGCATGACCTGTATATCCAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8808  
QY 601 TAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 656  
DB 8807 CAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8748  
QY 657 AGCAAGATCTCCATTAAGACACAAA 682  
DB 8747 AGTGAAGTATGCTCTCCAAAAAAA 8722

RESULT 13  
ABV84150/c  
ID ABV84150 standard; DNA; 11950 BP.  
XX  
XX ABV84150;  
AC  
XX  
XX  
DT 09-DEC-2002 (first entry)  
XX





Db 8867 GTGGCATGACCTGTAATCCAGCTACTCGGAGGCTGAGGACAGAGATCCCTGAAAC 8808  
Qy 601 TAGAGGTGAGGTGAGTAACCCGNG---ATGTCCTGCACTCCAGCCT-GGCACAG 656  
Db 8807 CAGGAGGTGAGGTGAGTGCAGAGATCAGCCATTGCACTCCAGCCTGGGCAACAG 8748  
Qy 657 AGCAAGACTCCATTAAGACACAAAA 682  
Db 8747 AGTGAGACTATGTCTCCAAAAAAA 8722

RESULT 14  
ADE43582/c  
ID ADE43582 standard; DNA; 128034 BP.  
XX ADE43582;  
AC  
XX  
XX 29-JAN-2004 (first entry)  
XX  
XX Polymorphic human IDE genomic sequence, SEQ ID 187.  
DE  
XX Neurodegenerative disease; uPA; SMC6; IDE; KNSL1; LIPA; TNFRSF6;  
KW Alzheimer's disease; neuroprotective; nootropic; gene therapy;  
XX Chromosome 10; gene; ds.  
XX  
OS Homo sapiens.

XX Key  
FH Location/Qualifiers  
FT 2456  
FT misc\_feature  
FT /tag= a  
FT /note= "There is a variation at this position"  
FT 3279  
FT misc\_feature  
FT /tag= b  
FT /note= "There is a variation at this position"  
FT 3407  
FT misc\_feature  
FT /tag= c  
FT /note= "There is a variation at this position"  
FT 17095  
FT misc\_feature  
FT /tag= d  
FT /note= "There is a variation at this position"  
FT 17242  
FT misc\_feature  
FT /tag= e  
FT /note= "There is a variation at this position"  
FT 33590  
FT misc\_feature  
FT /tag= f  
FT /note= "There is a variation at this position"  
FT 38903  
FT misc\_feature  
FT /tag= g  
FT /note= "There is a variation at this position"  
FT 42943  
FT misc\_feature  
FT /tag= h  
FT /note= "There is a variation at this position"  
FT 43391  
FT misc\_feature  
FT /tag= i  
FT /note= "There is a variation at this position"  
FT 45017  
FT misc\_feature  
FT /tag= j  
FT /note= "There is a variation at this position"  
FT 62498  
FT misc\_feature  
FT /tag= k  
FT /note= "There is a variation at this position"  
FT 68906  
FT misc\_feature  
FT /tag= l  
FT /note= "There is a variation at this position"  
FT 68973  
FT misc\_feature  
FT /tag= m  
FT /note= "There is a variation at this position"  
FT 69586  
FT misc\_feature  
FT /tag= n  
FT /note= "There is a variation at this position"  
FT 73772  
FT misc\_feature  
FT /tag= o

FT /note= "There is a variation at this position"  
FT 74084  
FT misc\_feature  
FT /tag= p  
FT /note= "There is a variation at this position"  
FT 83024  
FT misc\_feature  
FT /tag= q  
FT /note= "There is a variation at this position"  
FT 83104  
FT misc\_feature  
FT /tag= r  
FT /note= "There is a variation at this position"  
FT 89301  
FT misc\_feature  
FT /tag= s  
FT /note= "There is a variation at this position"  
FT 105060  
FT misc\_feature  
FT /tag= t  
FT /note= "There is a variation at this position"  
FT 107395  
FT misc\_feature  
FT /tag= u  
FT /note= "There is a variation at this position"  
FT 108489  
FT misc\_feature  
FT /tag= v  
FT /note= "There is a variation at this position"  
FT 111914  
FT misc\_feature  
FT /tag= w  
FT /note= "There is a variation at this position"  
FT 112114  
FT misc\_feature  
FT /tag= x  
FT /note= "There is a variation at this position"  
FT 113142  
FT misc\_feature  
FT /tag= y  
FT /note= "There is a variation at this position"  
FT 113591  
FT misc\_feature  
FT /tag= z  
FT /note= "There is a variation at this position"  
FT 114683  
FT misc\_feature  
FT /tag= aa  
FT /note= "There is a variation at this position"  
FT 116662  
FT misc\_feature  
FT /tag= ab  
FT /note= "There is a variation at this position"  
FT 117803  
FT misc\_feature  
FT /tag= ac  
FT /note= "There is a variation at this position"  
FT 124565  
FT misc\_feature  
FT /tag= ad  
FT /note= "There is a variation at this position"  
PN WO2003054143-A2.  
PD 03-JUL-2003.  
XX 25-OCT-2002; 2002WO-US034679.  
XX PF  
XX 25-OCT-2001; 2001US-0339525P.  
XX PR 08-NOV-2001; 2001US-0336929P.  
XX PR 08-NOV-2001; 2001US-0338010P.  
XX PR 09-NOV-2001; 2001US-0338363P.  
XX PR 04-DEC-2001; 2001US-0337052P.  
XX PR 28-MAR-2002; 2002US-0368919P.  
XX  
XX (NEUR-) NEUROGENETICS INC.  
XX (GEO) GEN HOSPITAL CORP.  
XX  
XX Becker KD, Velicelcbl G, Elliott KJ, Wang X, Tanzi RB, Bertram L;  
XX Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;  
XX MPI; 2003-559131/52.  
XX  
XX Determining a predisposition for or the occurrence of neurodegenerative  
XX disease, e.g. Alzheimer's disease by detecting in a target nucleic acid  
XX the presence or absence of an allelic variant of one or more polymorphic  
XX regions.



Db	81320	AAATTCAGTCAACTTTTACTGAGTTCCTTACCGAGGTCMAATTTGGGAAAGACAAATAT	81261
Qy	181	TCCTCTCAGCTGCTTTTCTGATATGATGAGATATAATGATGCTTTCAAGAGGAA	240
Db	81260	AGT-TCCTATTGGAAAACTCATATTATATGACACATGCTTATGCAAAATGTA	81202
Qy	241	TAAACCTATGAAAAGTTGAGGATAGTGTATGTAATGAAATTTCAACAATAG	300
Db	81201	TCAGTTATCTCATGTTAAACAGATGACAGACACTCCTCTATTATTACTGGTCTTG	81142
Qy	301	TAGCTGCTATTGAAAGTTTAAAGTTATTATTA-CACTATTATTAATAATTTAA	359
Db	81141	AAATTAACATGCTCTTTTAACTGTAAAGGTGACCAACACAGAAAAAGAGTTGA	81082
Qy	360	CTAATACACTTAATATTATTAAGAGCTTTGAATGGGCGACGCTGCTCTCCG	419
Db	81081	ATAGGTGAATGAGCCAAAGTATGAAACCACTCTTACGACGAGTGTGCTCATG	81022
Qy	420	TATCCCAACACTTTGGAGGCGCAAGTGGCGGATCACCTGAGTCAAGATTAA	479
Db	81021	TATCTCAACAAATTGGAGGCTGAGGCGGTGGATCATCTTGAAGTCAAGATTCA	80962
Qy	480	CAGCTGCGCAACATGTTGAAACCTGTCTCTAATAAAACGCAAAATTTAGCAG	539
Db	80961	CAGCTGCGCAATATGTTGAAACCTATCTCTAATAAAATTAAGCAAGTGT	80902
Qy	540	GGTGGATGACCTGTAGTCCCACTATCTCAGAGGTTGAGGAGAGAAATGCTTGA	599
Db	80901	GGTATGGGACCTGTATGATCCAGCCACTCGGAGGCTAGGAGGAGAAATCCTTGA	80842
Qy	600	CTAGAGGTGAGGTTGAGTAACCCGAGA--TGTCACTGCACTCAGGCT-GGCA	655
Db	80841	CCAGAGGCGCGAGCTGCGATGAGCCGAATTTGCACTTGCATCCAGCTGGGAA	80782
Qy	656	GAGCAAGACTCCATAAAGCAACAAAGCTTTGAAATTTGTAAATGAGTTGTAC	710
Db	80781	GAGCAAGACTGTCTCAAAAATATTAATAATTAAGAGAGAAATCACATCTCC	80727
RESULT 16			
ADH54059/c			
ID	ADH54059	standard; DNA; 128034 BP.	
XX	AC	ADH54059;	
XX	XX	25-MAR-2004 (first entry)	
DT	XX		
DE	XX	Human IDE gene DNA sequence SegID186.	
XX	XX	human; neurodegenerative disease; urokinase plasminogen activator; uPA;	
KM	XX	gamma-rynuclain; SNGs; insulin degrading enzyme; IDE;	
KM	XX	kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;	
KM	XX	tumour necrosis factor receptor Sfe; TNFRSF6; Alzheimer's disease; ds.	
OS	XX	Homo sapiens.	
XX	XX		
XX	XX	Key	Location/Qualifiers
FT	FT	variation	replace(2456,G)
FT	FT	variation	/tag= a
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(3279,C)
FT	FT	variation	/tag= b
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(3407,T)
FT	FT	variation	/tag= c
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(17095,G)
FT	FT	variation	/tag= d
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(17242,T)
FT	FT	variation	/tag= e
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(33590,G)
FT	FT	variation	/tag= f
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(38903,T)
FT	FT	variation	/tag= g
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(42943,C)
FT	FT	variation	/tag= h
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(43391,G)
FT	FT	variation	/tag= i
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(45017,T)
FT	FT	variation	/tag= j
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(62498,C)
FT	FT	variation	/tag= k
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(68906,G)
FT	FT	variation	/tag= l
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(68973,T)
FT	FT	variation	/tag= m
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(69586,C)
FT	FT	variation	/tag= n
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(73772,G)
FT	FT	variation	/tag= o
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(74084,T)
FT	FT	variation	/tag= p
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(83024,T)
FT	FT	variation	/tag= q
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(83104,T)
FT	FT	variation	/tag= r
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(89301,G)
FT	FT	variation	/tag= s
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(105060,G)
FT	FT	variation	/tag= t
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(107395,A)
FT	FT	variation	/tag= u
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(108489,T)
FT	FT	variation	/tag= v
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(111914,G)
FT	FT	variation	/tag= w
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(112114,A)
FT	FT	variation	/tag= x
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(113142,G)
FT	FT	variation	/tag= y
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(113591,G)
FT	FT	variation	/tag= z
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(114683,G)
FT	FT	variation	/tag= aa
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(116662,A)
FT	FT	variation	/tag= ab
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(117803,T)
FT	FT	variation	/tag= ac
FT	FT	variation	/standard name= "Single nucleotide polymorphism"
FT	FT	variation	replace(124565,T)

```
FT      /tag= ad
FT      /standard_name= "Single nucleotide polymorphism"
XX      US2003224380-A1.
XX      04-DEC-2003.
XX      25-OCT-2002; 2002US-00282174.
XX      25-OCT-2001; 2001US-0339525P.
XX      25-OCT-2001; 2001US-0348065P.
XX      02-NOV-2001; 2001US-0336983P.
XX      08-NOV-2001; 2001US-0336929P.
XX      08-NOV-2001; 2001US-0338010P.
XX      09-NOV-2001; 2001US-0338363P.
XX      04-DEC-2001; 2001US-0337052P.
XX      28-MAR-2002; 2002US-0368919P.
XX      (GENO ) GEN HOSPITAL CORP.
XX      Becker KD, Velicelabi G, Elliott KJ, Wang X, Tanzi RE;
XX      Bertiam L, Saunders AJ, Mullin KM, Sampson AJ;
XX      WPI; 2004-060538/06.
XX      Determining a predisposition for or the occurrence of neurodegenerative
XX      disease, particularly Alzheimer's disease, comprises determining the
XX      presence of a polymorphism in the uPA, SNGC, IDE, KNSL1, LIPA or TNFRSF6
XX      gene.
XX      Claim 22; SEQ ID NO 186; 205pp; English.
XX      This invention relates to a novel method of determining a predisposition
XX      for or the occurrence of neurodegenerative disease comprising detecting
XX      in a target nucleic acid obtained from the subject the presence of an
XX      allelic variant of polymorphic regions of human genes selected from
XX      urokinase plasminogen activator (uPA), gamma-synuclein (SNGC), insulin
XX      degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid
XX      lipase (LIPA) and tumor necrosis factor receptor SF6 (TNFRSF6). The
XX      method is useful in determining the presence or predisposition to a
XX      neurodegenerative disease, particularly Alzheimer's disease. The present
XX      sequence is that of the human IDE gene which is related to the invention.
XX      SEQ      Sequence 128034 BP; 34731 A; 25985 C; 26409 G; 40808 T; 0 U; 101 Other;
XX      Query Match      10.5%; Score 210.6; DB 12; Length 128034;
XX      Best Local Similarity 60.6%; Pred. No. 3.7e-34;
XX      Matches 433; Conservative 0; Mismatches 274; Indels 8; Gaps 5;
QY      1      TGATCCACGACGCTTGCCCTCCCAAGTGTGGATTACGGGGTGAGCCACGCGCTG 60
DB      81438      TGATCCACGCTGCTCGACCTCTCAAGTGTGGATTACGGGGTGAGCCACGCGCTG 81379
QY      61      GTCCGATGCTCTTATTATTGAGAGACAACATGCGGCTTAATCTGCTCTTATTGAC 120
DB      81378      GCTTAAGTCTTCTTAAGTAAGAAAAAGTCAAG--AAAGATTAGGCTTAAGTTAG 81321
QY      121      AGACTTGTATGATGATCAATCCCAATGCTGCCACTTACTGAACGGGCTTAATGACTTAG 180
DB      81320      AAATTCAGTCAACTTTTACTGAGTCTTACCAAGTCAAAATTTGCGGAAAGACAAATAT 81261
QY      181      TCTCTCTCAGCTGCTCTTTCGCAATGCTAGTGAAGTGAATATGATGCTTCAAGAGGAA 240
DB      81260      AGT-TCCTATTATGAAAAAAGCTCAATTAATTAATGAGACATGACTTAACCAAAATGTA 81202
QY      241      TAAACGTATGAAAAGTGTGAGAGTAGTGTGATATGAATAGATTTCAACAGTAG 300
DB      81201      TCAGTATATCTCATGTTAAAGATGAGAGACTACTCCTCTTATTACTGCTGCTTGG 81142
QY      301      TAGCTGCTATTTGAAGATTAAAGATTATTTATTA--CAACTATTATTAATTAATTTAAAA 359
DB      81141      AAATTAACATGGCTCTTTAAAGCTTAAGTGAACCAACAGAAAAAAGAGATTGAG 81082
```

```
QY      360      CTATATACCTTAATATTATTAAGAGCTTTGAATGGCCAGGCGCATAGCTCTGCTG 419
DB      81081      ATAGTGAATGAGCCAAATATGAAAACACCTCTCTAGCCAGGTGTGCTCATGCTG 81022
QY      420      TAATCCCAACACTTTGGAGGCGCAAGTGGGCGGATCACTGAGGTCAAGGATTAAAGC 479
DB      81021      TAATCTCAACAATTTGGAGGCGTGGGCGGATGATCACTGAGGTCAAGGATTCAAGAC 80962
QY      480      CAGCTGGCCAAACATGATGTAAGAACCTGTCTCTACTATTAAGCAAAATTAAGCCAGTGT 539
DB      80961      CAGCTGGCCAAATATGATGTAAGAACCTGTCTCTACTATTAAGCAAAATTAAGCCAGTGT 80902
QY      540      GGTGACATGACACTGTAGTCCCACTACTCAAGAGGTTGAGGAGAGAAATTCCTTGAAC 599
DB      80901      GGTAGTGGGCACTGTATATGACAGCCACTCGGAGGCTGAGGAGAGAGAAATTCCTTGAAC 80842
QY      600      CTAGAGGTGAGAGGTTGCAAGTACCCGAGA--TGCTCACTGACATCCAGGCT-GGCAACA 655
DB      80841      CCAGAGGCGCGAGGCTGCAAGTACCCGAGATTGTCCCAATGCACTCCAGGCTGGGCAACA 80782
QY      656      GAGCAAGACTCCATTAAGACAAACAAGCTTTGAATTTGTATATGAGTTGTAC 710
DB      80781      GAGCAAGACTCTGTCTCAAAAATATTAATAATTAAGAGAAATCATCTCC 80727
RESULT 17
ADHS4060/c
ID      ADHS4060 brandard; DNA; 128034 BP.
XX      ADHS4060;
XX      25-MAR-2004 (first entry)
XX      Human IDE gene variant DNA sequence SegID187.
XX      human; neurodegenerative disease; urokinase plasminogen activator; uPA;
XX      gamma-synuclein; SNGC; insulin degrading enzyme; IDE;
XX      kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;
XX      tumor necrosis factor receptor SF6; TNFRSF6; Alzheimer's disease; ds.
XX      OS      Homo sapiens.
XX      PN      US2003224380-A1.
XX      PD      04-DEC-2003.
XX      25-OCT-2002; 2002US-00282174.
XX      25-OCT-2001; 2001US-0339525P.
XX      25-OCT-2001; 2001US-0348065P.
XX      02-NOV-2001; 2001US-0336983P.
XX      08-NOV-2001; 2001US-0336929P.
XX      08-NOV-2001; 2001US-0338010P.
XX      09-NOV-2001; 2001US-0338363P.
XX      04-DEC-2001; 2001US-0337052P.
XX      28-MAR-2002; 2002US-0368919P.
XX      (GENO ) GEN HOSPITAL CORP.
XX      Becker KD, Velicelabi G, Elliott KJ, Wang X, Tanzi RE;
XX      Bertiam L, Saunders AJ, Mullin KM, Sampson AJ;
XX      WPI; 2004-060538/06.
XX      Determining a predisposition for or the occurrence of neurodegenerative
XX      disease, particularly Alzheimer's disease, comprises determining the
XX      presence of a polymorphism in the uPA, SNGC, IDE, KNSL1, LIPA or TNFRSF6
XX      gene.
XX      Claim 9; SEQ ID NO 187; 205pp; English.
XX      This invention relates to a novel method of determining a predisposition
XX      for or the occurrence of neurodegenerative disease comprising detecting
```

CC in a target nucleic acid obtained from the subject the presence of an  
CC allelic variant of polymorphic regions of human genes selected from  
CC urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin  
CC degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid  
CC lipase (LIPA) and tumor necrosis factor receptor SF6 (TNFRSF6). The  
CC method is useful in determining the presence or predisposition to a  
CC neurodegenerative disease, particularly Alzheimer's disease. The present  
CC sequence is that of the human IDE gene, with polymorphic sites  
CC represented by n, which is related to the invention.

XX  
SQ Sequence 128034 BP; 34726 A; 25977 C; 26400 G; 40799 T; 0 U; 132 Other;

Query Match 10.5%; Score 210.6; DB 12; Length 128034;  
Best Local Similarity 60.6%; Pred. No. 3.7e-34;

Matches 433; Conservative 0; Mismatches 274; Indels 8; Gaps 5;

QY 1 TGATCCACGAGCTTGGCTCCCAAGTGTGGATTACAGGCGTGAACCCACCGCTG 60

DB 81438 TGATCCACGAGCTTGGCTCCCAAGTGTGGATTACAGGCGTGAACCCACCGCTG 81379

QY 61 GTGGAATGCTTATTTATTTGAAGCAACATGGCGCTTAATCTGTCTTATTTGAC 120

DB 81378 GCTTAAGTGTCTTAAAGTAAAGAAAAGTCAAG-AAAAGTTAGCTTAAAGTTAG 81321

QY 121 AGACTTTGATGAGTCAAAATCCCAATGCTGCCACTTAAGTGAAGCGCTTAAATGACTAG 180

DB 81320 AAAATTCAGTCAACTTTTACGAGTCTTACCAAGTCAAAATTTGGGAAAAGCAAAATAT 81261

QY 181 TCTCTCTGAGCTGCTTTCTGCAATATGTAAGTGAATATGATGCTTTCAAGAGAA 240

DB 81260 AGT-TCTTATTTGGAAAACTCATATTTATGAGACATGACTATGCAAAATGTAA 81202

QY 241 TAAACCTAAGAAAAGTGTGAGATAGTGTGATATGAAATATGAAATTTCAACAAGTG 300

DB 81201 TGAATTAACTCATATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 81142

QY 301 TAGCTGCTATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 359

DB 81141 AATTAACATGCTCTTTAACTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 81082

QY 360 CTAAATCACTTAATTTAAAGCTTTGAAAATGGGCGCAGGCGAGTCTCTGCTG 419

DB 81081 ATAGGTGAATGAGCAAGTATGAAACCACTCTCTGCAAGGTGTGGTGCATGCTG 81022

QY 420 TAAATCCCAACATTTGGGAGGCGCAAGTGGCGGATCACTGAGTCAAGATTTAAAC 479

DB 81021 TAAATCCCAACATTTGGGAGGCGCGAGTGGATCACTGAGTCAAGATTTAAAC 80962

QY 480 CAGCTGGCGCAACATGTTGAAACCTGTCTCTACTAATAAAGCAAAATTAAGCAGGT 539

DB 80961 CAGCTGGCGCAACATGTTGAAACCTGTCTCTACTAATAAAGCAAAATTAAGCAGGT 80902

QY 540 GGTGGCATGACCTGTAGTCCCAACTACTCAGAGAGTTGAGGAGAGAAATGCTTGAAC 599

DB 80901 GGTGGCATGACCTGTAGTCCCAACTACTCAGAGAGTTGAGGAGAGAAATGCTTGAAC 80842

QY 600 CTAGAGAGTGTGAGCTTGGCTCCCAAGTGTGGATTACAGGCGTGAACCCACCGCTG 655

DB 80841 CTAGAGAGTGTGAGCTTGGCTCCCAAGTGTGGATTACAGGCGTGAACCCACCGCTG 80782

QY 656 GAGCAAGATCCATTAAGCAACAAAGCTTTGAATTTGTTAATGAGTTGATG 710

DB 80781 GAGCAAGATCCATTAAGCAACAAAGCTTTGAATTTGTTAATGAGTTGATG 80727

RESULT 18

ADBE43315  
ID ADBE43315 standard; DNA; 202100 BP.

XX ADBE43315;

XX 29-JAN-2004 (first entry)

XX

DE Human IDE/ KNSL1 genomic sequence, SEQ ID 484.

XX Neurodegenerative disease; uPA; SNCG; IDE; KNSL1; LIPA; TNFRSF6;

KM Alzheimer's disease; neuroprotective; neurotrophic; gene therapy;

XX Chromosome 10; gene; ds.

XX Homo sapiens.

PN W02003054143-A2.

XX 03-JUL-2003.

XX 25-OCT-2002; 2002WO-US034679.

XX 25-OCT-2001; 2001US-0339525P.

XX 08-NOV-2001; 2001US-0336829P.

XX 08-NOV-2001; 2001US-0338010P.

XX 09-NOV-2001; 2001US-0338363P.

XX 04-DEC-2001; 2001US-0337052P.

XX 28-MAR-2002; 2002US-0368919P.

PA (NEUR-) NEUROGENETICS INC.

PA (GEHO-) GEN HOSPITAL CORP.

PI Becker KD, Veliceli G, Elliott KJ, Wang X, Tanzi RE, Bertram L;

PI Saunders AJ, Mullin KM, Sampson AJ, Blacker DL;

XX WPI; 2003-559131/52.

XX Determining a predisposition for or the occurrence of neurodegenerative

PT disease, e.g. Alzheimer's disease by detecting in a target nucleic acid

PT the presence or absence of an allelic variant of one or more polymorphic

PT regions.

XX Claim 9; Page 769-823; 848pp; English.

XX The present invention relates to a method (M1) for determining a

CC predisposition for or the occurrence of neurodegenerative disease in a

CC subject. The method comprises detecting in a target nucleic acid obtained

CC from the subject the presence or absence of an allelic variant of one or

CC more polymorphic regions of one or more genes selected from uPA

CC (urokinase plasminogen activator), SNCG (gamma-synuclein), IDE (insulin-

CC degrading enzyme), KNSL1 (Kinesin-like protein 1), LIPA (lysosomal acid

CC lipase), and TNFRSF6 (Tumor Necrosis Factor Receptor-SF6), where the

CC presence of at least one of the allelic variant of one or more

CC polymorphic regions is indicative of a predisposition for or the

CC occurrence of neurodegenerative disease. The genes are all located on

CC chromosome 10. M1 is useful for determining a predisposition for or the

CC occurrence of, and for treating neurodegenerative disease, particularly

CC Alzheimer's disease.

XX

SQ Sequence 202100 BP; 60747 A; 41352 C; 41113 G; 58888 T; 0 U; 0 Other;

Query Match 10.5%; Score 210.6; DB 10; Length 202100;

Best Local Similarity 60.6%; Pred. No. 3.9e-34;

Matches 433; Conservative 0; Mismatches 274; Indels 8; Gaps 5;

QY 1 TGATCCACGAGCTTGGCTCCCAAGTGTGGATTACAGGCGTGAACCCACCGCTG 60

DB 42330 TGATCCACGAGCTTGGCTCCCAAGTGTGGATTACAGGCGTGAACCCACCGCTG 42389

QY 61 GTGGAATGCTTATTTATTTGAAGCAACATGGCGCTTAAATCTGTCTTATTTGAC 120

DB 42390 GCTTAAGTGTCTTAAAGTAAAGAAAAGTCAAG-AAAAGTTAGCTTAAAGTTAG 42447

QY 121 AGACTTTGATGAGTCAAAATCCCAATGCTGCCACTTAAGTGAAGCGCTTAAATGACTAG 180

DB 42448 AAAATTCAGTCAACTTTTACGAGTCTTACCAAGTCAAAATTTGGGAAAAGCAAAATAT 42507

QY 181 TCTCTCTGAGCTGCTTTCTGCAATATGTAAGTGAATATGATGCTTTCAAGAGAA 240

DB 42508 AGT-TCTTATTTGGAAAACTCATATTTAATATGAGACATGACTATGCAAAATGTAA 42566

QY 241 TAAACCTATGAAAGTGTGGAGATAGTGTGATGAAATAGAGATTCAACAGTNG 300  
DB 42567 TCAGTTAACTCATGTATTAACAGATGGCAGACACTCTCTCTATTATTTACTGATCTTTG 42626  
QY 301 TAGCTGCTATTTGAAGTTTAAAGTTATTATTA--CACTATTATTAATAATTTTAAAA 359  
DB 42627 AATTAACCTGCTCTTTAACTGTAAAGTGGACCAACACAGAAAAAGAGATTGAAG 42686  
QY 360 CTATATACACTTAAATTAATTAAGAGCTTTGAAATGGGCGCAGTACTCTGCTG 419  
DB 42687 ATAGGTGAATGAGCCAAAGTATGAAACCACTCTTAGCCAGGTGTGTGCTCATGCTTG 42746  
QY 420 TAATCCCAACACTTTGGGAGGCCAAGTGGCGGATCACTTAGTCAAGTAAAGAC 479  
DB 42747 TAATCTCAACAAATTTGGGAGGCGGTGAGCCGGTGAATCACTTAGGTCAAGATTCAAGAC 42806  
QY 480 CAGCTGGCCAACTGTGTGAAACCTGTCTCTACTAATAAAGCAAAATTTAGCCAGTGT 539  
DB 42807 CAGCTGGCCAAATATGTGTAACCCCTATCTTAATAAATACAAAATTTAGCCAGTGT 42866  
QY 540 GGTGGCATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 599  
DB 42867 GGTAGTGGCACTGTATGATGCCAGCACTCGGAGGCTGAGGAGAGAAATCTTTGAAC 42926  
QY 600 CTAGAGAGTGAAGTTGACGTAAACCCGAGA--TGTCACTGCACTCCAGCTT--GGCAACA 655  
DB 42927 CCAGAGAGGCCAGGCTGACGTAGCCAGCACTCGGAGTGTGCACTGCAAGCTTGGCAACA 42986  
QY 656 GAGCAAGACTCCATTAAGACAAACAGTTTGAATTTGTATTAAGATTGTAC 710  
DB 42987 GAGCAAGACTCTGTCTCAAAAATATTAATAAAGAGAAATCAATCTCC 43041

RESULT 19  
ADHS4357  
ID ADHS4357 standard; DNA; 202100 BP.  
XX  
AC ADHS4357;  
XX  
DT 25-MAR-2004 (first entry)  
XX  
DE Human IDE/KNSL1 gene DNA sequence Segid484.  
XX  
KW human; neurodegenerative disease; urokinase plasminogen activator; uPA;  
KW gamma-synuclein; SNCG; insulin degrading enzyme; IDE;  
KW kinesin-like protein 1; KNSL1; lysosomal acid lipase; LIPA;  
KW tumour necrosis factor receptor Sf6; TNFRSF6; Alzheimer's disease; ds.  
OS Homo sapiens.  
XX  
PN US2003224380-A1.  
XX  
PD 04-DEC-2003.  
XX  
PF 25-OCT-2002; 2002US-00282174.  
XX  
PR 25-OCT-2001; 2001US-0339525P.  
PR 25-OCT-2001; 2001US-0348065P.  
PR 02-NOV-2001; 2001US-0336983P.  
PR 08-NOV-2001; 2001US-0336929P.  
PR 08-NOV-2001; 2001US-0338010P.  
PR 09-NOV-2001; 2001US-0338363P.  
PR 04-DEC-2001; 2001US-0337052P.  
PR 28-MAR-2002; 2002US-0368919P.  
XX  
PA (GENO ) GEN HOSPITAL CORP.  
XX  
PI Becker KD, Veliceljebi G, Elliott KJ, Wang X, Tanzi RE;  
XX Berttram L, Saunders AJ, Mullin KM, Sampson AJ;  
DR MPI; 2004-060538/06.  
XX  
PT Determining a predisposition for or the occurrence of neurodegenerative

PT disease, particularly Alzheimer's disease, comprises determining the  
PT presence of a polymorphism in the uPA, SNCG, IDE, KNSL1, LIPA or TNFRSF6  
PT gene.  
XX  
XX  
PS Claim 9; SEQ ID NO 484; 205bp; English.  
XX  
CC This invention relates to a novel method of determining a predisposition  
CC for or the occurrence of neurodegenerative disease comprising detecting  
CC in a target nucleic acid obtained from the subject the presence of an  
CC allelic variant of polymorphic regions of human genes selected from  
CC urokinase plasminogen activator (uPA), gamma-synuclein (SNCG), insulin  
CC degrading enzyme (IDE), kinesin-like protein 1 (KNSL1), lysosomal acid  
CC lipase (LIPA) and tumour necrosis factor receptor Sf6 (TNFRSF6). The  
CC method is useful in determining the presence or predisposition to a  
CC neurodegenerative disease, particularly Alzheimer's disease. The present  
CC sequence is that of a region of human DNA, containing the human IDE and  
CC KNSL1 genes, which is related to the invention.  
XX  
SQ Sequence 202100 BP; 60747 A; 41352 C; 41113 G; 58888 T; 0 U; 0 Other;  
XX  
Query Match 10.5%; Score 210.6; DB 12; Length 202100;  
Best Local Similarity 60.6%; Pred. No. 3.9e-34;  
Matches 433; Conservative 0; Mismatches 274; Indels 8; Gaps 5;  
QY 1 TGAATCACCAGCTTGCTGCTCCCAAGTGTCTGGATTAAGCCGTGAGCACCAGCCCTG 60  
DB 42330 TGAATCACCAGCTTGCTGCTCCCAAGTGTCTGGATTAAGCCGTGAGCACCAGCCCTG 42389  
QY 61 GTGGAATGCTTTATTAATTTGAAGAGACAACATGGGCTTAATATCTGTCTTATTTGAC 120  
DB 42390 GCTTAAGTGTCTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 42447  
QY 121 AGACTTTGATGAGATCAAAATCCCAATGTCTCCCACTTACTGAAACGGCTTAAATGACTTAG 180  
DB 42448 AATATTCAGTCAACTTTTACTGATGTTCTTACCAAGGCAAAATCTGGGAAAGCAAAATAT 42507  
QY 181 TCTCTCAAGCTGTCTTCTGCAATATGATGATGATGATGATGATGATGATGATGATGATG 240  
DB 42508 AGT-TCTATTTTGAAGAAACCTATTAATTAATGACACATGACCTATGACAAATGTAA 42566  
QY 241 TAAACCTATGAAAGTGTGGAGATAGTGTGATGAAATAGAGATTCAACAGTNG 300  
DB 42567 TCAGTTAACTCATGTATTAACAGATGGCAGACACTCTCTCTATTATTTACTGATCTTTG 42626  
QY 301 TAGCTGCTATTTGAAGTTTAAAGTTATTATTA--CACTATTATTAATAATTTTAAAA 359  
DB 42627 AATTAACCTGCTCTTTAACTGTAAAGTGGACCAACACAGAAAAAGAGATTGAAG 42686  
QY 360 CTATATACACTTAAATTAATTAAGAGCTTTGAAATGGGCGCAGTACTCTGCTG 419  
DB 42687 ATAGGTGAATGAGCCAAAGTATGAAACCACTCTCTGAGGAGTGTGTGCTCATGCTTG 42746  
QY 420 TAATCCCAACACTTTGGGAGGCCAAGTGGCGGATCACTTAGTCAAGTAAAGAC 479  
DB 42747 TAATCTCAACAAATTTGGGAGGCGGTGAGCCGGTGAATCACTTAGGTCAAGATTCAAGAC 42806  
QY 480 CAGCTGGCCAACTGTGTGAAACCTGTCTCTACTAATAAAGCAAAATTTAGCCAGTGT 539  
DB 42807 CAGCTGGCCAAATATGTGTAACCCCTATCTTAATAAATACAAAATTTAGCCAGTGT 42866  
QY 540 GGTGGCATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 599  
DB 42867 GGTAGTGGCACTGTATGATGCCAGCACTCGGAGGCTGAGGAGAGAAATCTTTGAAC 42926  
QY 600 CTAGAGAGTGAAGTTGACGTAAACCCGAGA--TGTCACTGCACTCCAGCTT--GGCAACA 655  
DB 42927 CCAGAGAGGCCAGGCTGACGTAGCCAGCACTCGGAGTGTGCACTGCAAGCTTGGCAACA 42986  
QY 656 GAGCAAGACTCCATTAAGACAAACAGTTTGAATTTGTATTAAGATTGTAC 710  
DB 42987 GAGCAAGACTCTGTCTCAAAAATATTAATAAAGAGAAATCAATCTCC 43041

RESULT 20  
AEA61169\_2/c  
Continuation (3 of 4) of AEA61169 from base 200001 (Human CACNB2 gene genomic sequence S  
WP Sequence split into 4 fragments LOCUS AEA61169 Accession Aea61169  
WP Fragment Name Begin End  
WP AEA61169\_0 1 110000  
WP AEA61169\_1 100001 210000  
WP AEA61169\_2 200001 310000  
WP AEA61169\_3 300001 401433

Query Match 10.5%; Score 210.2; DB 14; Length 110000;  
Best Local Similarity 77.8%; Pred. No. 4,3e-34;  
Matches 267; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

QY 361 TATACCTTAATTAATTAAGAGCTTTGAATAGGCGCAGCTGAGTCTCTGCTGT 420  
DB 57128 TCAGGCCACATATTTTAAATTAAGCAAAATGCGCCAGACCAATAGCTCACAACCTGT 57069  
QY 421 AATCCCAACACTTTGGAGGCGCAAGGTGGCGGATCAGTCACTGAGGAGTTAAGACC 480  
DB 57068 AATCCCAACACTTTGGAGGCGCAAGGTGGCGGATCAGTCACTGAGGAGTTAAGACC 57009  
QY 481 AGCTTGCCCAACATGCTGAACCCCTGTCTTACTTAAACGCAAAAATTAGCCAGGTGTG 540  
DB 57008 AGCTTGCCCAACATGCTGAACCCCTGTCTTACTTAAACGCAAAAATTAGCCAGGTGTG 56949  
QY 541 GTGGCATGCACTGTAGTCCCACTACTCAGAGGTTAGAGGAGGAATTCCTTGAAACC 600  
DB 56948 GTGGCATGCACTGTAGTCCCACTACTCAGAGGTTAGAGGAGGAATTCCTTGAAACC 56889  
QY 601 TAGGAGGTGGAGGTGACAGTAACCCGAGA---TGTCATGCGACTCCAGCTGGCAACAG 657  
DB 56888 CTGGAGGTGGAGGTGACAGTAACCCGAGA---TGTCATGCGACTCCAGCTGGCAACAG 56829  
QY 658 GCAAGACTCCATTAAGACCAACAAAGCTTTGAATTTGTGTAA 700  
DB 56828 GCGAGACTCGTCTCAATCATCATCAATCAAAATGGAAATAGA 56786

RESULT 21  
ABX16034/c  
ID ABX16034 standard; DNA; 203654 BP.  
XX  
AC ABX16034;  
XX  
DT 04-APR-2003 (first entry)  
XX  
DE Human gene encoding calcium channel transporter family member.  
XX  
KM Human; ds; gene; chromosome 10; calcium channel transporter; cytosolic;  
KM liver; adrenal gland; tumorous nervous tissue; adult amygdala;  
KM brain meningioma; demis-draeh; adult brain; foetal brain; placenta;  
KM testis; kidney; SNP; single nucleotide polymorphism.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH variation replace(16802,G)  
FT FT /\*tag= a  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(17103,A)  
FT FT /\*tag= b  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(71782,C)  
FT FT /\*tag= c  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(102639,G)  
FT FT /\*tag= d  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(132707,C)  
FT FT /\*tag= e  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(140698,C)  
FT FT variation

FT FT /\*tag= f  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(140888,C)  
FT FT /\*tag= g  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(144278,Y)  
FT FT /\*tag= h  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(148851,C)  
FT FT /\*tag= i  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(159121,C)  
FT FT /\*tag= j  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(161499,G)  
FT FT /\*tag= k  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(161725,G)  
FT FT /\*tag= l  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(162859,T)  
FT FT /\*tag= m  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(162981,G)  
FT FT /\*tag= n  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(167468,R)  
FT FT /\*tag= o  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(167522,G)  
FT FT /\*tag= p  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(168351,G)  
FT FT /\*tag= q  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(168397,C)  
FT FT /\*tag= r  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(168534,T)  
FT FT /\*tag= s  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(168806,C)  
FT FT /\*tag= t  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(170017,T)  
FT FT /\*tag= u  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(171371,G)  
FT FT /\*tag= v  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(172828,T)  
FT FT /\*tag= w  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(174564,W)  
FT FT /\*tag= x  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(176912, .176914,AC)  
FT FT /\*tag= y  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(176912, .176914,AAC)  
FT FT /\*tag= z  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(182424,G)  
FT FT /\*tag= aa  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(183891, .183893,AT)  
FT FT /\*tag= ab  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(184320,A)  
FT FT /\*tag= ac  
FT FT /standard name= "Single nucleotide polymorphism"  
FT FT replace(184426,G)  
FT FT /\*tag= ad  
FT FT variation



FT	variation	/standard_name= "Single nucleotide polymorphism" replace(187085,G) /*tag= ae
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(187666,C) /*tag= af
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(187729,G) /*tag= ag
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(187734,T) /*tag= ah
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(187739,T) /*tag= ai
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(187740,A) /*tag= aj
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(188292,C) /*tag= ak
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(188742,G) /*tag= al
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(189033,T) /*tag= am
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(192341,Y) /*tag= an
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(192846,G) /*tag= ao
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(194004,C) /*tag= ap
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(194941,A) /*tag= aq
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(195301,C) /*tag= ar
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(200286, /*tag= as .200288,TT) /standard_name= "Single nucleotide polymorphism" replace(200322, /*tag= at .200324,AA) /standard_name= "Single nucleotide polymorphism" replace(200872, /*tag= au .200874,AC) /standard_name= "Single nucleotide polymorphism" replace(200872, /*tag= av .200874,AAC) /standard_name= "Single nucleotide polymorphism" replace(203298,T) /*tag= aw /standard_name= "Single nucleotide polymorphism"
XX		US2002142938-A1.
XX		03-OCT-2002.
XX		30-MAR-2001; 2001US-00820905.
XX		30-MAR-2001; 2001US-00820905.
XX		(YANC/) YAN C. (KERC/) KETCHUM K A. PA (DPRA/) DI FRANCESCO V. PA (BEAS/) BEASLEY E M.
PI	yan C, Ketchum KA, Di Francesco V, Beasley EM,	

XX		WP1: 2003-174139/17.	
DR	P-PSDB; ABG74119.		
XX	New peptides related to calcium channel transporter subfamily, useful for treating disorders associated with abnormal expression of the transporter in liver, adrenal gland, normal and tumorous nervous tissues, placenta, or kidney.		
PT			
PF			
XX			
PS	Claim 22; Fig 3; 193pp; English.		
CC	The invention relates to an isolated peptide (a member of the calcium channel transporter family) appearing as ABG74119 (or an allelic variant, orthologue or fragment of at least 10 contiguous amino acids), which is encoded by a nucleic acid molecule that hybridises under stringent conditions to the opposite strand of the cDNA and genomic DNA appearing as ABX16033 and ABX16034 (including their complement or a sequence 80% similar to them). Also included are antibodies, vectors, host cells, gene chips, transgenic non-human animals and identification of modulators and binding agents. The peptides are useful in substantial and specific assays related to functional information of the peptide sequences, to raise antibodies or to elicit immune response, as reagents in assays to		
Query Match	10.5%; Score 210.2; DB 10; Length 203654;		
Best Local Similarity	77.8%; Pred. No. 4,7e+34;		
Matches	267; Conservative 0; Mismatches 73; Indels 3; Gaps 1;		
Dn	361 TNATACACTTAAATTATAAAGACTTTGAATGGGCCAAGCGCAGTAGCTTCGTCTGT 57866 TCAGGGCCAATTTTTAAAATAAGCAAAATGGGGCCAGGCACAATAGCTCACAACTGT	 	420 574747
Oy	421 AATCCCACACTTTGGGAGGCCCAAAGTGSGCGCATCCTGAGAGTCAGAAGTTTAAGNC	 	480
Dd	57746 AATCCCCAGACTTTGGGAMCCAGAGGGGTGATCACTGTGTGTAGAGTTGCAGACC	 	57687
Oy	481 AGCCGTGGCAACATGTGNAAACCTGCTCTACTATAAAACGAAAAATTAGCAGGTGTG	 	540
Dd	57686 AGCTGCACCAACATGTGNAACCTGCTCTACTATAAAATTACAAAAATTAGCAGGTGTG	 	57627
Oy	541 GTGGCANTCAGCTGTAGTCCCAACTACTCAGAGAGTTGAAGSAGAGAAATYGCTTGAAAC	 	600
Dd	57626 GTGGCANTCAGTATMTATTCAGTACTGTGGAGGCGTGAAGGAGAGAAATCGCTGAACC	 	57567
Oy	601 TAGAGGTGAGGTGTCAGTAGTACCAG--TGTCATCTGCACTCCAGCTGTGCAACANA	 	657
Dd	57566 CTGGAGGTGAGGTGTGAGTAGAGCTGAGCACTGCACTCCAGGCTGTGCAACAGA	 	57507
Oy	658 GCAGAGCTCCATTAAGACACAAAAGTTGAAATGTGTAA	 	700
Dd	57506 GGAGACATCGCTCAATCAATCAATCAATCAAMAATGGAGATAGA	 	57464
RESULT 22			
ID ADQ20017	standard; DNA; 260160 BP.		
ADQ20017			
AC ADQ20017;			
DT 26-AUG-2004	(first entry)		
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 2837.			
KW soft tissue sarcoma; cytosstatic; gene therapy; vaccine; screening; human; ds.			
OS Homo sapiens.			
PN WO2004048938-A2.			
PD 10-JUN-2004.			
PF 26-NOV-2003; 2003WO-USO38193.			

XX	PR	26-NOV-2002; 2002US-0429739P.
XX	PA	(PROT-) PROTEIN DESIGN LABS INC.
XX	P1	Aziz N, Gineburg WM, Zlotnik A;
XX	XX	WPI, 2004-441208/41.
XX	PT	Early detection of soft tissue sarcoma comprises determining expression
XX	PR	of a gene in a first soft tissue sample and a normal soft tissue sample
XX	PT	and comparing the gene expression, also useful in treating soft tissue
XX	PS	sarcoma.
XX	XX	Example 2; SEQ ID NO 2837; 210pp; English.
XX	CC	The invention relates to a novel method for detecting soft tissue sarcoma
XX	CC	which comprises obtaining a first soft tissue sample from an individual
XX	CC	and a normal soft tissue sample from the same or different individual,
XX	CC	determining the expression of a gene in both samples and comparing the
XX	CC	expression of the gene in both soft tissue samples, where a higher level
XX	CC	of protein expression in the first soft tissue sample indicates the
XX	CC	presence of soft tissue sarcoma. The method of the invention has
XX	CC	cytostatic applications and may be useful for detecting soft tissue
XX	CC	sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX	CC	acid sequences may be useful in diagnostic and screening applications.
XX	CC	The current sequence is that of a human soft tissue sarcoma-upregulated
XX	CC	DNA of the invention. The current sequence is not shown within the
XX	CC	specification per se but was submitted in CD format by the inventor.
XX	XX	Sequence 260160 BP; 68455 A; 56280 C; 57380 G; 78045 T; 0 U; 0 Other;
XX	XX	Query Match 10.5%; Score 210.2; DB 12; Length 260160;
XX	XX	Best Local Similarity 74.4%; Pred. No. 4.9e-34;
XX	XX	Matches 279; Conservative 0; Mismatches 93; Indels 3; Gaps 1;
QY	311	TGAAGATTGAAGTATTATTTATTCACATTTTAAATTTTAAAACTAATACATT 370
DB	93728	TGGGCACCTTAGTAGATATCTGTTGAATATTTCTTTTTTTTAAATTAAAGAACTACAG 93787
QY	371	AAATTATTTAAAGACCTTTGAAATGGGCGACGCCAGTAGCTCTGCGTGAATCCCAACA 430
DB	93788	GAATTATGAGAAATGTAAAACTATGCGCTGGGTATGTGTGCTATGCTTAATCCAGCA 93847
QY	431	CTTTGGAGGCGCAAGGTGGGCGGATCCTAGGTCAGAGTTTAAAGCCAGCTGGCA 490
DB	93848	CTTTGGAGGCGTGAAGTGGGTGATCATCTGAGGTCCAGAAATTCAGACCAGCTGGCA 93907
QY	491	ACATGTGTAAACCTGTCTCTACTTAAAAACGCCAAAATTAGCCAGGTGTGTGGCATGCA 550
DB	93908	ACATGTGTAAACCTCTTCACTACTTAAAAAATTAACAAATTATGCCAGGTGTGTGGCATG 93967
QY	551	CCTGTAGTCCCACTACTCAGAGGTGAGGAGGAGGAATTTGTTAACCCTAGAGAGTGG 610
DB	93968	CTTGTAGTCCCACTTCTTCGGAGGCTGAGGAGGAGGAATTCGTTAAACCCGGAGGCGAG 94027
QY	611	AGTTTGCAGTAAACCCGAGAT---GTCACTGCACTCCAGCTGTGGCAACAGCAAGACTCC 667
DB	94028	AGTTTGCAGTAAACCCGAGATTTGGGCCACTGCACCTCCAGCTGTGGCAACAGCAAGACTCC 94087
QY	668	ATTAAGACCAACAAA 682
DB	94088	GTCTCAAAAAAAA 94102
XX	XX	RESULT 23
XX	XX	ADQ97721/C
XX	XX	ID ADQ97721 standard; DNA; 178024 BP.
XX	XX	ADQ97721;
XX	XX	AC
XX	DT	07-OCT-2004 (first entry)
XX	DB	Human cancer associated sequence HD10-033, SEQ ID 698.

XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2004060304-A2.  
XX  
XX 22-JUL-2004.  
XX  
XX 22-DEC-2003; 2003WO-US041389.  
XX  
XX 27-DEC-2002; 2002US-00330773.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX  
XX Morris DW, Malandro MS;  
XX  
XX WPI; 2004-543781/52.  
XX  
XX  
XX New isolated cancer associated nucleic acids comprising at least 10  
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating  
XX cancers such as leukemia and lymphoma.  
XX  
XX  
XX Claim 1; SEQ ID NO 698; 1999p; English.  
XX  
XX  
XX The present invention relates to cancer associated sequences (AD097025-  
XX CC AD098004). The sequences are useful for the diagnosis, prevention and/or  
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
XX CC data for this patent did not form part of the printed specification, but  
XX CC was obtained in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pat\_sequences.

Query Match	10.5%;	Score 209.2;	DB 12;	Length 178024;
Best Local Similarity	77.0%;	Pred. No. 7.5e-34;		
Matches 282;	Conservative 0;	Mismatches 78;	Indels 6;	Gaps 2;

  

Qy	320	AAGATTATTATTATACAACTATTTTAAATAAAATTTTAAAACTAATACACTTAATATTATTA	379
Db	50367	AAATCAATTTGGTTCGAATATTAAAGAAAGCTGCATTACAGAAACGTAAAGAAATTTTAA	50308
Qy	380	AAGAGCTTTGAAATGGCGCCAGGCGCAGTAGCTCTCTGCTGTATCCCAACACTTTGGGAG	439
Db	50307	TAAAACTTT--ACAGGCCGGGCGTGTGTGCTCACCGCTGTATCCAGCACCTTTGGGAG	50251
Qy	440	GCCAAAGGTGGCGGATCACCTGAGGTCAAGAGTTTAAAGCCAGCGTGGCCAACTGGTGA	499
Db	50250	GCTGAGGTGGGTGATCACTCGATGATCAGGAGTTCAAGACACAGCTTGGCCAACTGGTGA	50191
Qy	500	AACCTGTCTCTACTTAAAAACGCAAAATTTAGCAGAGTGGTGGCATCAGCTGTAGTC	559
Db	50190	AACCTGTCTCTACTTAAAAACAAAAATTTAGCATGATGGTGGCAGCTACCTGTAGTC	50111
Qy	560	CCAACTACTCAGAGGTTGAGGAGAGAAATTGCTTTGAACTTAGAGAGTGGAGTTGGCAG	619
Db	50130	CCAGCTACTTGGGAGGCTGAAGCAGAGAAATCGCTTGAACCCAGGAGGTGGAGTTGGCAG	50071
Qy	620	TAAACCGAAGT---GTCACTGCACCTCCAGCTGGCAACAGACAAAGACTCCATAAAGACA	676
Db	50070	TGAGATAGAGTGGCGCCATTGGCACTCCAGCCTGGTGACAGAACAGAACTGTCTCAAAAAA	50011
Qy	677	ACAAAA 682	
Db	50010	AAAAAA 50005	

  

RESULT 24
ADD71350/c
ID ADD71350 standard; DNA; 93390 BP.
AC ADD71350;
XX

DT 15-JAN-2004 (first entry)  
XX Glutamine:fructose-6-phosphate amide transferase 1 genomic sequence.  
DE diabetes; haplotype; polymorphism; diagnosis; renopathy; intron; ds;  
KW glutamine:fructose-6-phosphate amide transferase 1.  
XX  
OS Homo sapiens.  
PN WO2003023063-A1.  
PD 20-MAR-2003.  
PP 06-SEP-2002; 2002WO-JP009093.  
PR 07-SEP-2001; 2001JP-00271870.  
PR 28-MAR-2002; 2002JP-00090861.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Itakura M, Yasuno H, Watanabe I;  
XX  
DR WPI; 2003-313261/30.  
XX  
PT Judging relative onset risk of diabetes including type I or II diabetes  
PT and renopathy with or without type II diabetes accompanying, by detecting  
PT haplotype with gene polymorphism from human genomic DNA.  
XX  
PS Claim 1, SEQ ID NO 22; 157pp; Japanese.  
XX  
CC The invention relates to a method of judging the onset risk of diabetes  
CC comprising detecting a haplotype consisting of gene polymorphism at 1 or  
CC more positions selected from (a)-(h) from a specimen containing human  
CC genomic DNA supplied by a patient: (a) the nucleotide located at position  
CC 36 of the intron 1 on GFAT1 (glutamine:fructose-6-phosphate amide  
CC transferase 1) gene (nucleotide number 632 in sequence ADD71329; (b) the  
CC nucleotide located at position 7 of the intron 11 on GFAT1 gene  
CC (nucleotide number 266 in sequence ADD71330; (c) the nucleotide located  
CC at position -147 of the intron 12 on GFAT1 gene (nucleotide number 338 in  
CC sequence ADD71331; (d) the nucleotide located at positions 1853-1877 of  
CC the intron 8 on GFAT1 gene (nucleotide numbers 336-360 in sequence  
CC ADD71332; (e) the nucleotide located at positions 1988-2007 of the intron  
CC 12 on GFAT1 gene (nucleotide numbers 328-347 in sequence ADD71333; (f)  
CC the nucleotide located at position -11 to -22 of the intron 18 on GFAT1  
CC gene (nucleotide numbers 253-264 in sequence ADD71334; (g) the nucleotide  
CC located at positions 2632-2661 of the intron 3 on GFAT1 gene (nucleotide  
CC numbers 237-266 in sequence ADD71335; and (h) the nucleotide located at  
CC position 66 of the intron 18 on GFAT2 gene (nucleotide number 225 in  
CC sequence ADD71351). The method is useful for judging relative onset risk  
CC of diabetes including type I or II diabetes and renopathy with or without  
CC type II diabetes accompanying. This sequence represents the GFAT1 genomic  
CC sequence containing the polymorphisms shown above.  
XX  
SQ Sequence 93390 BP; 26851 A; 17344 C; 19969 G; 29217 T; 0 U; 9 Other;  
XX  
Query Match 10.4%; Score 208.8; DB 10; Length 93390;  
Best Local Similarity 75.7%; Pred. No. 8,4e-34;  
Matches 286; Conservative 0; Mismatches 87; Indels 5; Gaps 2;  
QY 308 TATTGAAGATTAAAGATTATTTTCACTATTATTAATTTTAAACATAATACA 367  
DB 32191 TCTCAAAAAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 32132  
QY 368 CTTAAATTATTAAGACCTTTGAAATGGCCGAGCGAGTAGCTCTGCTGTATCCCA 427  
DB 32131 CCAACATATATTAACAGAGGCTGATGGCCAGGCGAGGCTCAAGCCGTATATCCCA 32072  
QY 428 ACACTTTGGAGGCGAAGGTGGGGGATCCCTGAGTCGAGATTTAAGACGAGCTGG 487  
DB 32071 GCACTTTGGAGGCGGAGGCGGCGAGATCAC--GAGTCAGAGATGAGACCAATCTGG 32014  
QY 488 CCAACATGTGAACCTGTCTCTACTAAAAACGCAAAATTTAGCCAGGTGTGGTCAT 547

DB 32013 CCAACATGTGAACCCCTCTCTACTAATAAAATACAAAATTTAGTGGGTGTGGCAG 31954  
QY 548 GCACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGATTGCTTGAACCTAGAGG 607  
DB 31953 GCACCTGTATTCACAGCTACTCAGAGGGTGTAGGAGAGATTGCTTGAACCTGGAGG 31894  
QY 608 TGGAGTTGCAAGTAAACCCGAGAT--GTCACCTGCACTCCAGCCTGGCAACGAGCAAGC 664  
DB 31893 TGGAGGCTGCAAGTGAAGCCAAAGATCAGCCACTGCACTCAGCCTTGTGACAGAGTGAGC 31834  
QY 665 TCCATTAAGACACACAAA 682  
DB 31833 TCCATCTCAAAAAAAA 31816  
RESUT 25  
AD213446  
ID AD213446 standard; DNA; 243934 BP.  
XX  
AC AD213446;  
XX  
DT 16-JUN-2005 (first entry)  
XX  
DE Human cancer-associated genomic DNA #82.  
XX  
KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;  
KW cytostatic; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2005031001-A2.  
XX  
PD 07-APR-2005.  
XX  
PF 23-SEP-2004; 2004WO-US011617.  
XX  
PR 23-SEP-2003; 2003US-00669920.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Morris DW, Malandro MS;  
XX  
DR WPI; 2005-273395/28.  
XX  
PT Nucleic acid array useful for detecting cancer associated nucleic acid,  
PT comprises two or more nucleic acid probes.  
XX  
PS Disclosure; SEQ ID NO 966; 198pp; English.  
XX  
CC The invention relates to a nucleic acid array for detecting a cancer  
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
CC The invention also relates to a peptide array comprising two or more  
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
CC that binds to a polypeptide, an isolated antibody or its fragment which  
CC binds to a polypeptide, which is prepared by immunizing a host animal  
CC with a composition comprising the polypeptide or its antigen binding  
CC fragment and collecting cells from the host expressing antibodies against  
CC the antigen or its antigen binding fragment, a composition comprising the  
CC antibody and a carrier, a method of screening for anticancer activity, a  
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
CC method of treating cancer and a method of inhibiting expression of a CA  
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
CC nucleic acids. The antibody is useful for detecting the presence or  
CC absence of cancer cells in an individual which involves contacting cells  
CC from the individual with the antibody and detecting a complex of a CA  
CC protein from the cancer cells and the antibody, where the detection of  
CC the complex correlates with the presence of cancer cells in the  
CC individual. The composition is useful for inhibiting growth of cancer  
CC cells in an individual or for delivering a therapeutic agent to cancer  
CC cells in an individual. The invention is also useful for diagnosing  
CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
CC a cell. This sequence represents human cancer-associated genomic DNA of  
CC the invention.

XX Sequence 243934 BP; 66213 A; 49129 C; 53448 G; 74301 T; 0 U; 843 Other;  
SQ  
Query Match 10.4%; Score 208.8; DB 14; Length 243934;  
Best Local Similarity 77.0%; Pred. No. 9.6e-34;  
Matches 269; Conservative 0; Mismatches 77; Indels 3; Gaps 1;  
QY 359 ACTAATACACTTAAATTTAAAGAGCTTTGAATGAGGCGAGGAGCTAGCTCCCTGCT 418  
DB 74796 AGATTGTGGCAGACATTTTGAATAATGACAAGATGAGCGGCGCGGAGCTTCACACCT 74855  
QY 419 GTAATCCCAACACTTTGGAGGCCAAGTGGGCGGATCACCTGAGGTCAAGATTAAAGA 478  
DB 74856 GTATCCAGACACTTTGGAGGCCAAGTGGGCGGATCACCTGAGGTCCGAGTTCAAGA 74915  
QY 479 CCGAGCTGGCCAACTGCTGTAACCTCTGCTCTAATAAAACGAAAAATTAGCCAGGTG 538  
DB 74916 CCGAGCTGGCCAACTGCTGTAACCTCTGCTCTAATAAAATCAAAAAATTAGCTGGGTG 74975  
QY 539 TGGTGGCATGCACTGTAAGTCCCACTACTCAGAGGTGAGGAGGAGAAATTGCTTGA 598  
DB 74976 TGGTGGCATGCACTGTAAGTCTGCTGCTAGCTAGAGGCTGAGGAGGAGAAATTGCTTGA 75035  
QY 599 CCTAGGAGGTGAGGTTGACAGTAACCGAGA---TGTCACTGCACTCCAGCTGGCAACA 655  
DB 75036 CCTGAGAGGTGAGGTTGACAGTAACCTGAGATGCTGCCATTGCCAGCTGGCAAC 75095  
QY 656 GAGCAAGACTCCATPAAAGACACAAAGCTTTGAATTTGTAAATGA 703  
DB 75096 AAGAGTGAAGCTCCATTTCAAAAAAGAAAAAAGAAATGACAAAGTGA 75143  
RESULT 26  
ID ADC86824/c  
XX ADC86824 standard; DNA; 17588 BP.  
AC  
XX  
ADBC86824;  
XX  
DT 01-JUN-2004 (first entry)  
XX  
DE Human GPCR gene SEQ ID NO:1277.  
XX  
ds: gene: human; GPCR;  
XX  
KM guanosine triphosphate-binding protein coupled receptor; gene therapy.  
XX  
OS Homo sapiens.  
XX  
XX  
PN EP1270724-A2.  
XX  
PD 02-JUN-2003.  
XX  
PF 18-JUN-2002; 2002EP-00013517.  
XX  
PR 18-JUN-2001; 2001JP-00246789.  
XX  
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATOR.  
XX  
PI Suwa M, Aseai K, Akiyama Y, Aburatani H;  
XX  
XX WPI: 2003-315783/31.  
DR P-PSDB; ADC86825.  
XX  
PT New polynucleotide, useful for preparing a composition for treating a  
PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.  
XX  
PS Claim 1; SEQ ID NO 1277; 28pp; English.  
XX  
XX The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient

CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The  
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
CC invention.  
XX  
SQ Sequence 17588 BP; 4287 A; 4481 C; 4353 G; 4267 T; 0 U; 200 Other;  
Query Match 10.4%; Score 208.6; DB 10; Length 17588;  
Best Local Similarity 80.6%; Pred. No. 7.3e-34;  
Matches 257; Conservative 0; Mismatches 59; Indels 3; Gaps 1;  
QY 390 AAATGGGCGAGGCGAGTACTCTGCTGTAATCCCAACACTTTGGAGGCCAAGGTG 449  
DB 15131 AAATTGGCTGGGCGAGTACTCAACCTGTATCCCAACACTTTGGAGGCCAAGGTG 15072  
QY 450 GCGGATCAGCTGAGTGCAGAGTTTAAAGACCAAGCTGGCCCAATGATGTAACCTGCT 509  
DB 15071 GCGGATTAAGCTGAGTGCAGAGTTCAAGACCAAGTGGCCCAATGATGTAACCTGCT 15012  
QY 510 CTACTAAAAACGAAAAATTAGCCAGTGTGTTGAGTGCATGCTGTAATCCCAACTATC 569  
DB 15011 CTACTAAAAATPACAAAAATTAGCCAGTGTGTTGAGTGCATGCTGTAATCCCAACTAT 14952  
QY 570 AGGAGGTGAGGAGGAGAAATTGCTTGAACCTAGAGGTGAGAGTTGACAGTAACCGAGA 629  
DB 14951 GGGAGCTGAGGAGGAGAAATTGCTTGAACCTAGAGGTGAGAGTTGACAGTAACCGAGA 14892  
QY 630 T---GTCACTGCACTCCAGCTGGCAACAGAGCAAGCTTCATPAAAGACAAAGGCTT 686  
DB 14891 TCTCGGCATGCACTCCAGCTGGCAACAGAGCAAGCTTCATPAAATPAAATA 14832  
QY 687 TGAATTTGTGTAATGACT 705  
DB 14831 AATPAAATPAAATPAAATPAAAT 14813  
RESULT 27  
ID ADO48536/c  
XX ADO48536 standard; DNA; 63266 BP.  
AC  
XX  
ADDO48536;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE Human neuropilin 1 (NRP1) genomic DNA sequence #2.  
XX  
XX  
KM human; melanoma; single nucleotide polymorphism; SNP; neuropilin 1; NRP1;  
XX  
KM mannose receptor C type 2; MRC2; gene; ds.  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
XX  
FH variation 32929  
FT /tag= a  
FT /note= "single nucleotide polymorphism"  
FT variation 34983  
FT /tag= b  
FT /note= "single nucleotide polymorphism"  
XX  
PN MO2004044163-A2.  
XX  
PD 27-MAY-2004.  
XX  
PF 06-NOV-2003; 2003WO-US035876.  
XX  
PR 06-NOV-2002; 2002US-0424475P.  
XX  
PR 23-JUL-2003; 2003US-0489703P.  
XX  
XX (SEQU-) SEQUENOM INC.  
XX  
XX Roth RB, Nelson MR, Braun A, Kammerer SM;  
PI  
XX WPI: 2004-411720/38.  
DR

```
XX Identifying a subject at risk of melanoma, useful for treating melanoma,
PT comprises detecting the presence or absence of one or more polymorphic
PT variations associated with melanoma in a nucleic acid sample from a
PT subject.
XX
XX Claim 23; SEQ ID NO 2; 176bp; English.
XX
XX The invention comprises a method for identifying a subject at risk of
XX melanoma. The invention involves detecting the presence or absence of one
XX or more polymorphic variations associated with melanoma in the neupilin
XX 1 (NRPI) or mannose receptor C type 2 (MRC2) genes. The method of the
XX invention is useful for identifying subjects at risk and treating
XX melanoma. The present nucleic acid represents the genomic DNA sequence
XX for human NRPI.
XX
XX Sequence 63266 BP; 18633 A; 14176 C; 13204 G; 17251 T; 0 U; 2 Other;
SQ
Query Match 10.4%; Score 208.6; DB 12; Length 63266;
Best Local Similarity 77.1%; Pred. No. 8.7e-34;
Matches 280; Conservative 0; Mismatches 79; Indels 4; Gaps 2;
QY 326 TATTTATTTCACTATTTAATAAATTTTAAACCTAATCACTTAATTTAAAGAGC 385
DB 34581 TATTTATTTCAACCTGTTTCAGCTGATGATTCAGGCTGCTTTAAAAAATAATC 34522
QY 386 TTTGAATGGGCGAGGCGAGTAGCTCTGCTGTAATCCCAACATTTGGAGGCCAAG 445
DB 34521 ACAGAAAGAGGCCAGGTGCAATGGCTCATGCTGTATTCACCACTTTGGAGGCCAAG 34462
QY 446 GTGGGCGGATCACCTGAGTCAAGAGTTTAAAGACAGCCCTGGCCAAATGTTGAAACCT 505
DB 34461 GTGGGTGATCACCTGAGTCAAGAGTTTAAAGACAGCCCTGGCCAAATGTTGAAACCT 34402
QY 506 GTCTCTACTTAAACCGAAATAATTAGCCAGGTGTGGCATGCACTGTATGCCCAACT 565
DB 34401 GTTTCTACTTAAATAATCAAAATAATTAGAGGAGTGTGGGACGCTCTCTGGATCCAGCT 34342
QY 566 ACTCAGAGGTGAGGAGGAGAGATGCTGAACCTTGAAGGTGGAGGTGGAGTGAACCC 625
DB 34341 ACTTGGAGGCTGAGGAGGAGAGATGCTGAACCTTGAAGGTGGAGGTGGAGTGAACCC 34282
QY 626 GAGA--TGTCACTGCACTCCAGCCT-GGCAACGAGCAAGACTCCATAAAGACAA 681
DB 34281 GAGATTGTGCACTTGTACTCCAGCTGGGAGCAAAAGCAAGCTCCATCTCAAAAACAA 34222
QY 682 AGC 684
DB 34221 ACC 34219
RESULT 28
ADX80721/c
ID ADX80721 standard; DNA; 63266 BP.
XX
XX ADX80721;
XX
XX 05-MAY-2005 (first entry)
XX
XX Human nidogen 2 (NID2) genomic DNA.
XX
XX melanoma; DNA polymorphism; SNP detection; cyrostatic; gene therapy; SNP;
XX single nucleotide polymorphism; gene; de.
XX
XX Homo sapiens.
XX
XX Key location/Qualifiers
XX variation 32929
XX FT /tag= a
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT 34983
XX FT /*tag= b
XX FT /standard_name= "Single nucleotide polymorphism"
```

```
XX
XX WO2005017176-A2.
XX
XX 24-FEB-2005.
XX
XX 05-MAY-2004; 2004WO-US014238.
XX
XX 23-JUL-2003; 2003US-0489703P.
XX
XX 06-NOV-2003; 2003US-00703789.
XX
XX 06-NOV-2003; 2003US-00703817.
XX
XX 06-NOV-2003; 2003US-00704513.
XX
XX (SEQU-) SEQUENOM INC.
XX
XX Roth RB, Nelson MR, Kammerer SM, Braun A, Hoyal-Wrightson CR;
XX
XX WPI: 2005-182387/19.
XX
XX P-PSDB; ADX80738.
XX
XX
XX Identifying a subject at risk of melanoma by detecting presence or
XX absence of a polymorphic variation associated with melanoma, where the
XX presence of polymorphic variations is indicative of the subject being at
XX risk of melanoma.
XX
XX Example 21; SEQ ID NO 2; 418bp; English.
XX
XX The invention relates to a novel method for identifying a subject at risk
XX of melanoma. The method comprises detecting the presence or absence of a
XX polymorphic variation associated with melanoma, where the presence of the
XX one or more polymorphic variations is indicative of the subject being at
XX risk of melanoma. The invention further comprises: a method for
XX identifying a polymorphic variation associated with melanoma proximal to
XX an incident polymorphic variation associated with melanoma; an isolated
XX nucleic acid which comprises a portion of or all of a nucleotide sequence
XX comprising fully defined 68400-21330 base pairs sequences (SEQ ID NO. 3,
XX 4, 5, 6, and/or 7) given in the specification, and comprises one or more
XX polymorphic variations; an oligonucleotide comprising a nucleotide
XX sequence complementary to a portion of the nucleotide sequence above,
XX where the 3' end of the oligonucleotide is adjacent to a polymorphic
XX variation; a microarray comprising the isolated nucleic acid linked to a
XX solid support; an isolated polypeptide encoded by the isolated nucleic
XX acid sequence; genotyping a nucleic acid; a method for identifying a
XX candidate molecule that modulates cell proliferation; treating melanoma
XX in a subject; and treating melanoma in a subject or preventing melanoma
XX in a subject. The methods and sequences have cyrostatic activity. The
XX polymucleotides may be used in gene therapy. The methods are useful for
XX identifying a subject at risk of melanoma, treating melanoma in a
XX subject, or preventing melanoma in a subject. This polymucleotide
XX sequence represents the polymorphic variation containing human nidogen 2
XX (NID2) genomic DNA of the invention.
XX
XX Sequence 63266 BP; 18633 A; 14176 C; 13204 G; 17251 T; 0 U; 2 Other;
SQ
Query Match 10.4%; Score 208.6; DB 14; Length 63266;
Best Local Similarity 77.1%; Pred. No. 8.7e-34;
Matches 280; Conservative 0; Mismatches 79; Indels 4; Gaps 2;
QY 326 TATTTATTTCACTATTTAATAAATTTTAAACCTAATCACTTAATTTAAAGAGC 385
DB 34581 TATTTATTTCAACCTGTTTCAGCTGATGATTCAGGCTGCTTTAAAAAATAATC 34522
QY 386 TTTGAATGGGCGAGGCGAGTAGCTCTGCTGTAATCCCAACATTTGGAGGCCAAG 445
DB 34521 ACAGAAAGAGGCCAGGTGCAATGGCTCATGCTGTATTCACCACTTTGGAGGCCAAG 34462
QY 446 GTGGGCGGATCACCTGAGTCAAGAGTTTAAAGACAGCCCTGGCCAAATGTTGAAACCT 505
DB 34461 GTGGGTGATCACCTGAGTCAAGAGTTTAAAGACAGCCCTGGCCAAATGTTGAAACCT 34402
QY 506 GTCTCTACTTAAACCGAAATAATTAGCCAGGTGTGGCATGCACTGTATGCCCAACT 565
DB 34401 GTTTCTACTTAAATAATCAAAATAATTAGAGGAGTGTGGGACGCTCTCTGGATCCAGCT 34342
```

OY	566	ACTCAGGAGGCTTGAAGGGAGGAGAATTGCCTTGAACCTTAGAGAGGTTCGATGTAACC	625
Db	34341	ACTTGCGAGGCTTGAGGCAAGAAATCATCTGAACCTGGGAAGTGAAGTTGCAGTGAACC	342822
OY	626	GAGA---TGTCACTGCATCCAGCTC--GGCAAGCAGAGCAAGACTCCATTAAGACAACAA	681
Db	34281	GAGATGTGGCATTTGTACTCCAGCTCGGGGGAACAAGACTCCATCTCAAAAACAA	342222
OY	682	AGC 684	
Db	34221	ACC 34219	
 RESULT 29			
ID	ADQI7329	standard; DNA; 101685 BP.	
XX	ADQI7329;		
XX	ADQI7329;		
XX	26-AUG-2004	(first entry)	
DT	26-AUG-2004	(first entry)	
DE	Human soft tissue sarcoma-upregulated DNA - SEQ ID 146.		
XX	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;		
KW	ds.		
XX	Homo sapiens.		
OS	WO2004048938-A2.		
PN	10-JUN-2004.		
PD	26-NOV-2003; 2003WC-USO38193.		
PF	26-NOV-2002; 2002US-0429739P.		
PR	(PROT-) PROTEIN DESIGN LABS INC.		
PA	Aziz N, Ginsburg WM, Zlotnick A;		
PI	WPJ; 2004-441208/41.		
DR	Early detection of soft tissue sarcoma comprises determining expression		
XX	of a gene in a first soft tissue sample and a normal soft tissue sample		
PT	and comparing the gene expression, also useful in treating soft tissue		
PT	sarcoma.		
XX			
PS	Example 2; SEQ ID NO 146; 210bp; English.		
XX	The invention relates to a novel method for detecting soft tissue sarcoma		
CC	which comprises obtaining a first soft tissue sample from an individual		
CC	and a normal soft tissue sample from the same or different individual,		
CC	determining the expression of a gene in both samples and comparing the		
CC	expression of the gene in both soft tissue samples, where a higher level		
CC	of protein expression in the first soft tissue sample indicates the		
CC	presence of soft tissue sarcoma. The method of the invention has		
CC	cycrostatic applications and may be useful for detecting soft tissue		
CC	sarcoma, possibly via gene therapy or vaccine production. The nucleic		
CC	acid sequences may be useful in diagnostic and screening applications.		
CC	The current sequence is that of a human soft tissue sarcoma-upregulated		
CC	DNA of the invention. The current sequence is not shown within the		
CC	specification per se but was submitted in CD format by the inventor.		
XX			
SQ	Sequence 101685 BP; 27434 A; 22468 C; 23225 G; 28558 T; 0 U; 0 Other;		
 Query Match            10.4%; Score 208.6; DB 12; Length 101685; Best Local Similarity   58.8%; Pred. No. 9.3e-34; Matches   433; Conservative   0; Mismatches   274; Indels   30; Gaps   3;			
OY	1	TGATTCACCGAGCCCTGGGCTCCCAAABTGTGGGATTATACAGGCGTAGCACACGCCCTG 60	
Db	51671	TGATTCACCTGCTCGGCTCCCAAABTGTGGGATTATACAGGCGATATAGCCACGCCCTT 51730	

QY	6	GTGCAATGCTTATATTTTGAAGAGACAACATGGGCGTTAAATCGCTTCTATTTGAC	120
Db	51731	GCCAGGAGATTCCTTAAACTTTAGTAGATATAAAAGCTCCCTCGAGAGTTGTATAA	51786
QY	121	AGACTTGATGAGAGTCAAAATCCCAATGCTGCACACTTACTGAACGGCGCTTAATGACTTAG	180
Db	51787	ACAAATTCCTGGGCTTCAATCCAGAGATTTGACTTAGTAGGTGTTCCAGTACTTACT	51846
QY	181	TCTCTCTCAGCTGTCTTTCGTGATATGTAAAGTGAAATATGATAGGCTTTCAAGAGAA	240
Db	51847	TCTAAAGAAATACCCCTAGGTGATTTTGGGGTATCTCTGGACCAACTTTGA--GATACC	51904
QY	241	TAAACCTATGAAAAAGTGTGAGATATGCTGTTGATATGAAATATGAGATTTCAACAATGAG	300
Db	51905	TGACCCGATGGGAGATGTATAGACATTTAATTTTACTTTTGTGTGACATACATCTGCT	51964
QY	301	TAGCTGCTATTTGAAGATTTTAAAGTTATTTATATACATATTTAATATAATTTTAAAAAC	360
Db	51965	TACTTTATATAATTAATCTCAGATAGTGTTTAATAATATTTTAAATTAATTAATAAATTT	52024
QY	361	TAATACACTTAAATTTATTAAGAAGCTTTGAAATAGGCGCAGGCGAGTAGCTCTCGCTGT	420
Db	52025	TAGAA-----TTAGACGACGAGGCGTGTAGCTCAGACCTGT	52060
QY	421	AATCCCAACACTTTGGGAGGCGCAAGGTGGCGGATCAGCTGAGGTGAGGATTTTAAAGCC	480
Db	52061	AATCTTAGCACTTTGGGAGGCCAGGATAGGCAGATCAGCTGAGGTGAGGATTTTGAAGCC	52120
QY	481	AGCTTGCGCAACATGTGTAAACCTGTCTTACTTAAAAACGAAAAATTTAGCCAGGTGTG	540
Db	52121	AGCTTGCGCAACATGTGTAAACCTGTCTTCTTAAAAATTAACAAAATTTAGCTGGGTGTG	52180
QY	541	GTGCGATGCACTGTGATGTCACACTACTCAGAGAGGTGAGGAGAGAAATTCCTTGAAGCC	600
Db	52181	GTGCGGTGTCAATGTATTCACACTACCCAGAGAGGTGTGAGGAGAGAAATTCCTTGAAGCC	52240
QY	601	TAGAGAGTGAAGGTGTCAGTAAACCCAGATGTCACTGCACTCAGGCTTGCAACAGAGCA	660
Db	52241	CGGAGAGCGAGAGGTGTCAGTGAAGAGATCGCATCTGCGACGCTTAGGCAACAGAGCA	52300
QY	661	AGACTTCATTAAGACAAACAAGCTTTGAAATTTGTGTAAATGATGATTTGATCTTCAAT	720
Db	52301	AGACTTCATCTCAAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT	52360
QY	721	TTTAAAGAAATTCATCTTT 737	
Db	52361	AGAAGGATACCTGTGT 52377	

RESULT 30	
AA104768	
ID	AA104768 standard; DNA; 8553 BP.
XX	
XX	
AC	AA104768;
XX	
XX	
DT	21-NOV-2001 (first entry)
XX	
DE	Human reproductive system related antigen DNA SEQ ID NO: 7456.
XX	
XX	
KM	Human; reproductive system related antigen; reproductive system disorder;
XX	
XX	cancer; gene therapy; ds.
OS	Homo sapiens.
XX	
PN	WO200155320-A2.
XX	
PD	02-AUG-2001.
XX	
XX	
PE	17-JAN-2001; 2001WO-US001339.
XX	
XX	
PR	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-020515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 14-AUG-2000; 2000US-0226279P.  
PR 18-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227099P.  
PR 23-AUG-2000; 2000US-0227099P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234224P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251889P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.

Disclosure, SEQ ID NO 7456; 1297bp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a



CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention

XX Sequence 8553 BP; 2253 A; 1971 C; 1958 G; 2372 T; 0 U; 0 Other;

Query Match 10.4%; Score 208.4; DB 4; Length 8553;  
Best Local Similarity 84.8%; Pred. No. 7,2e-34;  
Matches 246; Conservative 0; Mismatches 41; Indels 3; Gaps 1;

QY 396 GCCAGGCGAGTAGTCTCTGCTGTAATCCCACTTTGGAGGCGAAGTGGCGGAT 455  
DB 541 GCCGGGCGCTGTGGCTCAGCGCTGTATCCACAGCTTTGGAGGCGGCGGTGGAT 600  
QY 456 CACCTGAGGTCAAGAGTTTAAGCCAGCGCTGGCCAACTGTGTCTTACTA 515  
DB 601 CACCTGAGGTCAAGAGTTTGAAGCCAGCGCTGGCCAACTGTGTCTTACTA 660  
QY 516 AAAACGCAAAATTTAGCCAGGTGTGGGATGACCTGTAGTCCCACTACTCAGGAG 575  
DB 661 AAAATTCAAAAATTTAGCCGCGGTGTGGGCGCGCTGTATCCAGCTACTCAGGAG 720  
QY 576 TTGAGGAGAGAGATTGCTTGAACCTAGAGGTGAGGTTGCAGTAACCGAGA--TGT 632  
DB 721 CTGAGGAGAGAGATGCTTTGAACCTGAGAGCGAGGTTGCAGTACGAGCATGTGT 780  
QY 633 CACTGACTCCAGCTCGGCGACAGAGCAAGACTCCATAAGACACAAA 682  
DB 781 CACTGACTCCAGCTCGGCGACAGAGCAAGACTCCATAAGACACAAA 830

## RESULT 31

ABL97671  
ID ABL97671 standard; DNA; 8553 BP.

XX ABL97671;

XX 21-JUN-2002 (first entry)

DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2323.

KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;

KW reproductive system disorder; urinary system disorder; gene therapy;

KW cardiovascular disorder; respiratory disorder; neurological disorder;

KW gastrointestinal disease; infection; cytostatic; gene; ds.

OS Homo sapiens.

XX Homo sapiens.

XX WO200155317-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001329.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-020515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225475P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232403P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.

PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250311P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 05-DAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483232/52.  
XX  
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful  
XX PT for preventing, diagnosing and/or treating testicular cancer.  
XX  
XX  
XX  
XX Disclosure; SEQ ID NO 2323; 766pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 973  
XX CC human testicular antigens, and fragments of their genomic sequences. The  
XX CC sequences can be used in the treatment of cardiovascular, urinary system,  
XX CC reproductive system, immune, respiratory, neurological and  
XX CC gastrointestinal disorders, infections, and particularly cancer,  
XX CC especially testicular cancers. The present sequence is a DNA encoding a  
XX CC protein fragment of the invention  
XX  
XX Sequence 8553 BP; 2252 A; 1971 C; 1958 G; 2372 T; 0 U; 0 Other;

Query Match 10.4%; Score 208.4; DB 4; Length 8553;  
Best Local Similarity 84.8%; Pred. No. 7.2e-34;  
Matches 246; Conservative 0; Mismatches 41; Indels 3; Gaps 1;  
396 GCCAGGCGAGTCTGCTGCTGTATCCCAACACTTGGAGGCCAAGTGGCGGAT 455  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 541 GCCGCGCCTGTGCTCAGCCCTGTATCCAGCATTTGGAGGCCGAGCGGTGAT 600  
Qy 456 CACCTGAGTCCAGAGTTTAAAGACCGAGCTGGCCAACTGGTGAACCTGTCTACTA 515  
Db 601 CACCTGAGTCCAGAGTTTGAAGACCGAGTGGCCAACTGGTGAACCTGTCTACTA 660  
Qy 516 AAAACGCAAAAATTAGCCAGGTGTGTGGCATTCGACTGTGTGCCCACTACACGAGG 575  
Db 661 AAAATACAAAATTATAGCCGCGGTGTGTGGCATTCGACTGTGTGCCCACTACAGAGG 720  
Qy 576 TTGAGGAGGAGGAATTTGTTGAACCTAGAGGTGAGGTGAGTACCCGAGA---TGT 632  
Db 721 CTGAGGCGAGGAATCGCTTGAACCTGAGCGAGGTGTGAGTGAAGCCGACATCGTGT 780  
Qy 633 CACTGCACTCCAGCTTGCGCAACAGACGAAGTCCATTAAGACAAACAAA 682  
Db 781 CACTGCACTCCAGCTTGCGCGACAGCGAGCTCATTCAAAAA 830  
RESULT 32  
ACN44462  
ID ACN44462 strand; DNA; 51928 BP.  
XX  
XX ACN44462;  
AC  
XX  
XX 18-NOV-2004 (first entry)  
XX  
XX Human genomic sequence hCG38450.  
DE  
XX  
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO2003073826-A2.  
PN  
XX  
XX 12-SEP-2003.  
PD  
XX  
XX 28-FEB-2003; 2003WO-US006235.  
PF  
XX  
XX 01-MAR-2002; 2002US-00087192.  
PR  
XX  
XX (SAGR-) SAGRES DISCOVERY.  
PA  
XX  
XX Morris DW;  
PI  
XX  
XX WPI; 2003-328604/31.  
DR  
XX  
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX PT comprises a nucleotide sequence.  
XX  
XX  
XX Claim 1; SEQ ID NO 922; 0pp; English.  
PS  
XX  
XX The present invention relates to novel DNA and protein sequences which  
XX CC are associated with carcinomas. The sequences are useful for: (i) for  
XX CC screening drug candidates; (ii) for screening of bioactive agent capable  
XX CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
XX CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biologic;  
XX CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX CC carcinoma including lymphoma. The present sequence is one such CA coding  
XX CC sequence. Note: This patent is an equivalent to basic patent  
XX CC US2002182586A1, for which no sequence data was published  
XX  
XX Sequence 51928 BP; 12432 A; 9654 C; 9978 G; 13530 T; 0 U; 6334 Other;

Query Match 10.4%; Score 208.4; DB 11; Length 51928;  
Best Local Similarity 81.2%; Pred. No. 9.3e-34;  
Matches 255; Conservative 0; Mismatches 56; Indels 3; Gaps 1;

OY	372	AATATTAAAGGCTTTGAAATGGGCCAGGCGCAGTGTCTCGCCGTAAATCCAAAC	431
Db	27729	AATGAAAAAAATTTTGGCCTTTGGCCAGCGCGGTGCTCAGCTGTATCTCAGCAC	27788
OY	432	TTTGGAGAGCCAGAGTGGCGGATCACTGAGTCAAGATTAAAGCCAGCCTGGCCAA	491
Db	27789	TTTGGAGAGCCAAAGTGGGTGGATCATCTGAGGTTAGAGTTTGAACAGCCTGGCCAA	27848
OY	492	CATGGTGAACACCCCTGTCTTACTTAAAAACGCAAAATTTAGCCAGGTGTGGCATGCAC	551
Db	27849	CATGGTGAACACCCCATATCTACTTAAAAATTCAAAAATTTAGCGGGGTGTGTATGCAC	27908
OY	552	CTGTAGTCCCAACTACTACAGGAGTTGAGGAGAGAAATTGCTTGAACCTAGGAGGTGGA	611
Db	27909	CTGTATATCCACCTACTCTCCGAGGCTGAGGCTGAGATTTGCTTGAACCCAGAGGGGGA	27968
OY	612	GGTTGCAAGTAACCCGAGAT--GTCACTGCACCTTCAGCCTGCAGACAGACAGACTCCA	668
Db	27969	GGTTGCAAGTGAAGCTGAAGATCAAGGCCATTGCACCTCGGCTGGGTAACGAGAACTGTA	28028
OY	669	TTAAGACACACAAA 662	
Db	28029	TTTCAAAAAAAA 28042	
RESULT 33			
ADZ12979	ID	ADZ12979 standard; DNA; 52124 BP.	
XX	AC	ADZ12979;	
XX	DT	16-JUN-2005 (first entry)	
XX	DE	Human cancer-associated genomic DNA #44.	
XX	KW	Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;	
XX	KM	Cytostatic; gene; db.	
OS	XX	Homo sapiens.	
PN	XX	WO2005031001-A2.	
PD	XX	07-APR-2005.	
PF	XX	23-SEP-2004; 2004WO-US031617.	
PR	XX	23-SEP-2003; 2003US-00669920.	
PA	XX	(CHIR ) CHIRON CORP.	
PI	XX	Morris DW, Malandro MS;	
PT	XX	WPI; 2005-273395/28.	
PS	XX		
XX	XX	Disclosure; SEQ ID NO 499; 198bp; English.	
<p>The invention relates to a nucleic acid array for detecting a cancer associated (CA) nucleic acid, comprising two or more nucleic acid probes. The invention also relates to a peptide array comprising two or more isolated polypeptides encoded by a CA nucleic acid sequence, a compound that binds to a polypeptide, an isolated antibody or its fragment which binds to a polypeptide, which is prepared by immunizing a host animal with a composition comprising the polypeptide or its antigen binding fragment and collecting cells from the host expressing antibodies against the antigen or its antigen binding fragment, a composition comprising the antibody and a carrier, a method of screening for anticancer activity, a method of detecting a CA nucleic acid, a method of diagnosing cancer, a method of treating cancer and a method of inhibiting expression of a CA nucleic acid in a cell. The CA nucleic acids are useful for detecting CA nucleic acids. The antibody is useful for detecting the presence or</p>			

[illegible]

PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216680P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0226279P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236377P.  
PR 29-SEP-2000; 2000US-0236377P.  
PR 29-SEP-2000; 2000US-0236377P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239393P.  
PR 13-OCT-2000; 2000US-0239397P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-46570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
XX Disclosure; SEQ ID NO 8396; 1297bp + Sequence Listing; English.  
XX  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX  
XX Sequence 27154 BP; 7884 A; 5528 C; 6156 G; 7586 T; 0 U; 0 Other;  
SQ



Db 232396 TGAAGAGCAGATCTGTTCTCAAAAAAAAAA 232426  
RESULT 36  
AAK68575  
ID AAK68575 standard; DNA; 32145 BP.  
XX  
AC AAK68575;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23387.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218230P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228287P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0231415P.

PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0233398P.  
PR 14-SEP-2000; 2000US-0233399P.  
PR 14-SEP-2000; 2000US-0233400P.  
PR 14-SEP-2000; 2000US-0233401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234979P.  
PR 25-SEP-2000; 2000US-0234989P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.

17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX MPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclosure; SEQ ID NO 23387; 3071bp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX SQ Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 U; 0 Other;  
Query Match 10.4%; Score 208; DB 4; Length 32145;  
Best Local Similarity 83.8%; Pred. No. 1.1e-33;  
Matches 248; Conservative 0; Mismatches 45; Indels 3; Gaps 1;  
QY 390 AATATGGGCGAGCGCAGTAGCTCTGCTGTAATCCCAACTTTGGAGGCCAAGGTGG 449  
DB 25596 ATAGAGGCGCAGGTGCGGTGGCTCACACCTGCAATCCACACACTTTGGAGGCCAAGGTGG 25655  
QY 450 GGGGATCACTGAGGTTCAGAGGTTTAAGACCAACCCCTGGCCACATGTGTAACCCCTGTCT 509  
DB 25656 GGAGATCACTGAGGTTCAGAGGTTTAAGACCAACCCCTGGCCACATGTGTAACCCCTGTCT 25715  
QY 510 CTACTTAAATGCAAAATTTAGCCAGGTGTTGGTGGCATGCACCTGTAGTCCCAACTACATC 569  
DB 25716 CTACTTAAATGCAAAATTTAGCCAGGTGTTGGTGGCATGCACCTGTAGTCCCAACTACATC 25775  
QY 570 AGGAGTTGAGGAGGAGGAATTTGCTTGAACCTAGAGGTTGAGGTTGACAGTAAACCGAG- 628  
DB 25776 AGGAGTTGAGGAGGAGGAATTTGCTTGAACCTAGAGGTTGAGGTTGACAGTAAACCGAG- 25835  
QY 629 -ATGTCATGCACTTCAGCTCAGCTGGGCAACGAGCAATCTCCATTAAGACAACCAAA 662  
DB 25836 TATGTCATGCTTCAGCTCAGCTGGGCAATCTCCATTAAGACAATCTCTCTCAGAAAAAAA 25891

RESULT 37  
AAK68491

ID AAK68491 standard; DNA; 32145 BP.  
XX  
XX AAK68491;  
AC  
XX  
DT 06-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23303.  
DE Human immune/haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157182-A2.  
XX  
XX  
XX 09-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-022513P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226779P.  
XX 22-AUG-2000; 2000US-0226881P.  
XX 22-AUG-2000; 2000US-0226888P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 01-SEP-2000; 2000US-0229509P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.  
XX 08-SEP-2000; 2000US-0231243P.  
XX 08-SEP-2000; 2000US-0231244P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 08-SEP-2000; 2000US-0231414P.  
XX 08-SEP-2000; 2000US-0232080P.  
XX 08-SEP-2000; 2000US-0232081P.  
XX 12-SEP-2000; 2000US-0231968P.  
XX 14-SEP-2000; 2000US-0232397P.  
XX 14-SEP-2000; 2000US-0232398P.



PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246533P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclosure; SEQ ID NO 23303; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients' own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 U; 0 Other;  
SQ  
Query Match 10.4%; Score 208; DB 4; Length 32145;  
Best Local Similarity 83.8%; Pred. No. 1,1e-33;  
Matches 248; Conservative 0; Mismatches 45; Indels 3; Gaps 1;  
QY 390 AATGGGCGAGGCGAGTCTCTGTAATCCCACTTTGGAGGCCAAGTGG 449  
DB 25596 ATRAGAGCCAGGTGGCGGTGCTCACACTGCAATCCCACTTTGGAGGCCAAGTGG 25655  
QY 450 GCGGATCACCTGAGGTCTGAGGTTTAAGACCAAGCTGGCCAAATGTTAAACCTGTCT 509  
DB 25656 GCAGATCACCTGAGGTCTGAGGTTTGAAGACCAAGCTGGCCAAATGTTAAACCTGTCT 25715  
QY 510 CTACTAAAAACGCAAAAATTTAGCCAGGTGTGTGGATGATCCTTGATGCCCACTACTC 569  
DB 25716 CTACTAAAAATGCAAAAATTTAGCCAGGTGTGTGGATGATCCTTGATGCCCACTACTC 25775  
QY 570 AGGAGTTGAGGAGGAGGAATGCTGAACCTAGAGGTGAGAGTTGACGTAACCCGAG- 628  
DB 25776 AGGAGGCTAGGAGGAGGAATGCTGAACCTGAGGAGGTGAGAGTTGACGTAACCCGAGA 25835  
QY 629 --ATGTCACTGCACTCCAGCTTGGCAACAGAGCAAGCTTCATTAAGACCAACAAA 682  
DB 25836 TCATGCCACTGTACTCCAGCTTGGCAATAGAGCTAATTTCTGTCTCAGAAAAAAA 25891  
RESULT 38  
AADS8977 ID AAD58977 standard; DNA; 83450 BP.  
AC AADS8977;  
XX  
XX 18-DEC-2003 (first entry)

XX	Human phosphatase genomic DNA.	
DE		
XX	Human: phosphatase protein; therapeutic; diagnostic; immune response;	
KW	pharmacogenomic; tissue typing; gene therapy; chromosome 15; enzyme;	
KW	cancer; gene; ds.	
XX		
XX	Homo sapiens.	
OS		
XX		
XX		
Key	Location/Qualifiers	
FT	variation	replace(85, T)
FT		/*tag= a
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(1614, G)
FT		/*tag= b
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(2648, G)
FT		/*tag= c
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(6773, G)
FT		/*tag= d
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(8144, A)
FT		/*tag= e
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(9488, C)
FT		/*tag= f
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(9730, G)
FT		/*tag= g
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(10317, C)
FT		/*tag= h
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(10391, C)
FT		/*tag= i
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(10682, A)
FT		/*tag= j
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(11632, .11634, TT)
FT		/*tag= k
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(12054, A)
FT		/*tag= l
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(12938, G)
FT		/*tag= m
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(13719, A)
FT		/*tag= n
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(14332, .14334, TT)
FT		/*tag= o
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(14753, G)
FT		/*tag= p
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(14830, .14832, AT)
FT		/*tag= q
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(14831, A)
FT		/*tag= r
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(15157, A)
FT		/*tag= s
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(16242, A)
FT		/*tag= t
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(16698, G)
FT		/*tag= u
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"

---

FT	variation	replace(17152, G)
FT		/*tag= v
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(17623, .17623, TA)
FT		/*tag= w
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(17624, T)
FT		/*tag= x
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(18267, .18269, AA)
FT		/*tag= y
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(18786, G)
FT		/*tag= z
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(18812, .18814, CG)
FT		/*tag= aa
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(18913, A)
FT		/*tag= ab
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(22061, G)
FT		/*tag= ac
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(22205, T)
FT		/*tag= ad
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(22527, A)
FT		/*tag= ae
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(22876, G)
FT		/*tag= af
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(23351, A)
FT		/*tag= ag
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(23821, G)
FT		/*tag= ah
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(25487, T)
FT		/*tag= ai
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(30858, C)
FT		/*tag= aj
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(31014, C)
FT		/*tag= ak
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(31457, T)
FT		/*tag= al
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(35341, G, A)
FT		/*tag= am
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(35346, G)
FT		/*tag= an
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(36164, A)
FT		/*tag= ao
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(36620, G)
FT		/*tag= ap
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(37074, G)
FT		/*tag= aq
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(37545, .37547, TA)
FT		/*tag= ar
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(37546, T)
FT		/*tag= as
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		replace(38189, .38191, AA)

```
FT      /tag= ac
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(38708, A)
FT      /tag= au
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(38734, .38736, CG)
FT      /tag= av
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(38735, A)
FT      /tag= aw
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(41983, G)
FT      /tag= ax
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(42127, T)
FT      /tag= ay
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(42449, A)
FT      /tag= az
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(42798, G)
FT      /tag= ba
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(43273, A)
FT      /tag= bb
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(43743, G)
FT      /tag= bc
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(45407, C)
FT      /tag= bd
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(45985, T)
FT      /tag= be
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(46198, G)
FT      /tag= bf
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(46792, .46194, AG)
FT      /tag= bg
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(46922, .46924, AT)
FT      /tag= bh
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(47851, G)
FT      /tag= bi
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      /tag= bj
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(48875, T)
FT      /tag= bj
```

Query Match 10.44; Score 208; DB 10; Length 83450;  
Best Local Similarity 77.8%; Pred. No. 1.2e-33;  
Matches 277; Conservative 0; Mismatches 75; Indels 4; Gaps 2;

```
QY      361 TAATACCTTAATTAATTAAGAGCTTTGAATGGCGAGCGAGTAAGCTCCGCTGT 420
DB      60529 TTATGAATTTCCCGATGAAAAAGTACAAATGAGCGCGGTGGCTCATGCTCT 60588
QY      421 AATCCCAACCTTTGGAGGCGCAAGTGGCGGATCACTAGGTCAGAGTTTAAGACC 480
DB      60589 AATCCCAACCTTTGGAGGCGCAAGGCGGTGGCTCATGAGTCAGAGTTTAAGACC 60648
QY      481 AGCTGGGCAACATGGTGAACCTGTCTCTACTAATAAAAGCAAAATTAAGCCAGGTGTG 540
DB      60649 AGCTGGGCAACATGGTGAACCTGTCTCTCTAATAAATTAAGCAAGCATG 60708
QY      541 GTGGCAGTCACTGTAGTCCCACTACTCAGAGGTTTGAAGGAGGAGGAGTGTGAACC 600
DB      60709 GTGGCAGTCACTGTAGTCCCACTACTCAGAGGTTTGAAGGAGGAGGAGTGTGAACC 60768
QY      601 TAGGAGGTGAGGTTTGAAGTAAACCGAGAT--GTCACATGCACTCCAGCTT-GGCAAG 656
DB      60769 TGGAGGCGAAGGTTTGAAGTAAACCGACCTGCACTGCAAGCTTGGGTGAATG 60828
```

```
QY      657 AGCAAGACTCCATAAAGCAAAAGCTTGAATGTGTAAATGAGTTGACT 712
DB      60829 AGCAAGTCTCATCTCAAAAAAAATTTACAAAGTGGCGCCGCTGTGTGAGCT 60884

RESULT 39
ADG88582
ID      ADG88582 standard; DNA; 83450 BP.
XX
AC      ADG88582;
XX
DT      11-MAR-2004 (first entry)
XX
DE      Human phosphatase gene.
XX
KW      Human; phosphatase; immune response; cancer; gene therapy; chromosome 15;
XX      ds.
XX
OS      Homo sapiens.
XX
XX      Location/Qualifiers
FT      Key
FT      PH      variation
FT      replace(85,T)
FT      /tag= a
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(1614,G)
FT      /tag= b
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(12648,G)
FT      /tag= c
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(6773,G)
FT      /tag= d
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(8144,G)
FT      /tag= e
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(9488,C)
FT      /tag= f
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(9730,G)
FT      /tag= g
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(10317,C)
FT      /tag= h
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(10391,C)
FT      /tag= i
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      /tag= j
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(11633,TT)
FT      /tag= k
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      replace(12054,A)
FT      /tag= l
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      /tag= m
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      /tag= n
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      /tag= o
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      /tag= p
FT      /standard name= "Single nucleotide polymorphism (SNP)"
FT      /tag= q
FT      /standard name= "Single nucleotide polymorphism (SNP)"
```

```
FT variation replace(15157,A)
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(16242,A)
FT /tag= s
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(16698,G)
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(17152,G)
FT /tag= u
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(17623, .17625, TA)
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(17624,T)
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18267, .18269, AA)
FT /tag= x
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18786,G)
FT /tag= y
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18812, .18814, CG)
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(18813,A)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(22061,G)
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(22205,T)
FT /tag= ac
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(22527,A)
FT /tag= ad
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(22876,G)
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(23351,A)
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(23821,G)
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(25487,T)
FT /tag= ah
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(30858,C)
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(31014,C)
FT /tag= aj
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(31437,T)
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(35341,GA)
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(35346,G)
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(36164,A)
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(36620,G)
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(37074,G)
FT /standard_name= "Single nucleotide polymorphism (SNP)"

FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(37545, .37547, TA)
FT /tag= aq
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(37546,T)
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(38189, .38191, AA)
FT /tag= as
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(38734, .38736, CG)
FT /tag= at
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(38735,A)
FT /tag= au
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(41983,G)
FT /tag= av
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(42127,T)
FT /tag= aw
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(42449,A)
FT /tag= ax
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(42798,G)
FT /tag= ay
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(43273,A)
FT /tag= az
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(43743,G)
FT /tag= ba
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(45407,C)
FT /tag= bb
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(45985,T)
FT /tag= bc
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(46198,G)
FT /tag= bd
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(46792, .46794, AA)
FT /tag= be
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(46922, .46924, AT)
FT /tag= bf
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(47851,G)
FT /tag= bg
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(48675,T)
FT /tag= bh
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(50844,C)
FT /tag= bi
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace(51267,T)
FT /tag= bj
FT /standard_name= "Single nucleotide polymorphism (SNP)"

Query Match 10.4%; Score 208; DB 10; Length 83450;
Best Local Similarity 77.8%; Pred. No. 1.2e-33;
Matches 277; Conservative 0; Mismatches 75; Indels 4; Gaps 2;

OY 361 TATACACTTAATATATTAAGAGCTTGAATGGCCAGGCGCAGTACCTCTCCCTCT 420
DB 60529 TTATGAATTTCCCTATATAAAAATGAGCCAGGCGTGTGCTATCCCTGT 60588
OY 421 AATCCCAACCTTGGAGGCGCAAGTGGGCGATCACCTGAGTCAGAGTTTAAGACC 480
||||| ||||||| ||||||| ||| ||||||| ||||||| ||||||| ||||||| |||||||
```

Db	60589	AATCCAGACTTTGGGAGGCCAAGCGGGTGCTCACTGAGGATTCAGACC	60648	FT	variation	replace(11633,-)
Qy	481	AGCCGCGCAATGATGTAACCCGTCTCTCTAATAAAGCAAAATTTAGCAGGTGG	540	FT	variation	/tag= m
Db	60649	AGCCTGGCAACATGGTAACCTGTCTCTGTAATAAATTAGCCAGCATG	60708	FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
Qy	541	GTGGATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATTCGTAACC	600	FT	variation	/tag= n
Db	60709	GTGGACGACACTGTAGTCCAGCTACTCAGGAGGCTGAGGAGAGATTCAGTAACC	60768	FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
Qy	601	TAGAGGTGGAGGTTGCGCTAACCGAGAT---GTCTCTGCTACTCCAGCTT-GGCACAG	656	FT	variation	/tag= o
Db	60769	TGGGAGGCGAGAGTTGCTGATGATCAAGATCAAGCCACTGCGAGCTGGGTGATAG	60828	FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
Qy	657	ACCAAGACTCCCTAAGACAAACAGCTTGAAATTTGTAATGATTGTAACCT	712	FT	variation	/tag= p
Db	60829	ACCAAGTCTCCATCTCAAAAAAAATTTACAAAGTGGCGCGGTGGTAGCT	60884	FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
RESULT 40				FT	variation	/tag= q
AEBA7870				FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
ID	AEBA7870	standard; DNA; 83450 BP.		FT	variation	/tag= r
XX	AEBA7870;			FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
XX	06-OCT-2005	(first entry)		FT	variation	/tag= s
DE	Human novel splice form phosphatase genomic DNA.			FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
XX	splice variant; cytoslatic; gene therapy; cancer; neoplasm; phosphatase;			FT	variation	/tag= t
KW	ds; gene; chromosome 15; SNP; single nucleotide polymorphism.			FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
XX	Homo sapiens.			FT	variation	/tag= u
XX	Key	Location/Qualifiers		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/tag= a		FT	variation	/tag= v
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	replace(1614,G)		FT	variation	/tag= w
FT	variation	/tag= b		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/tag= x
FT	variation	replace(1698,G)		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/tag= c		FT	variation	/tag= y
FT	CDS	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		2075..72181		FT	variation	/tag= z
FT		/tag= d		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT		/product= "Human novel splice form phosphatase protein"		FT	variation	/tag= aa
FT		/note= "this CDS contains numerous introns, not defined within the specification"		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	replace(2648,G)		FT	variation	/tag= ab
FT	variation	/tag= e		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/tag= ac
FT	variation	replace(6773,G)		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/tag= f		FT	variation	/tag= ad
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	replace(8144,G)		FT	variation	/tag= ae
FT	variation	/tag= g		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/tag= af
FT	variation	replace(9488,C)		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/tag= h		FT	variation	/tag= ag
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	replace(9730,G)		FT	variation	/tag= ah
FT	variation	/tag= i		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/tag= ai
FT	variation	replace(10317,C)		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/tag= j		FT	variation	/tag= aj
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	replace(10391,C)		FT	variation	/tag= ak
FT	variation	/tag= k		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/tag= al
FT	variation	replace(10682,A)		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	/tag= l		FT	variation	/tag= al
FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"		FT	variation	/standard_name= "Single nucleotide polymorphism (SNP)"
FT	variation	replace(35341,G/A)		FT	variation	/tag= am

```
FT /tag= ak
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(35346,G)
FT /tag= al
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(36164,A)
FT /tag= am
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(36620,G)
FT /tag= an
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(37074,G)
FT /tag= ao
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(37546,T/-)
FT /tag= ap
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(38190,-)
FT /tag= aq
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(38708,A)
FT /tag= ar
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(38735,A/-)
FT /tag= as
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(41983,A)
FT /tag= at
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(42127,T)
FT /tag= au
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(42449,A)
FT /tag= av
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(42798,G)
FT /tag= aw
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(43273,A)
FT /tag= ax
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(43743,G)
FT /tag= ay
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(45407,C)
FT /tag= az
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(45985,T)
FT /tag= ba
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(46198,G)
FT /tag= bb
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(46793,-)
FT /tag= bc
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(46923,-)
FT /tag= bd
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(47851,G)
FT /tag= be
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(48875,T)
FT /tag= bf
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(50844,C)
FT /tag= bg
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(51267,T)
FT /tag= bh
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(54073,G)
FT /tag= bi
```

```
FT /standard name= "Single nucleotide polymorphism (SNP)"
FT replace(54206,G)
FT variation
Query Match 10.4%; Score 208; DB 14; Length 83450;
Best Local Similarity 77.8%; Pred. No. 1.2e-33;
Matches 277; Conservative 0; Mismatches 75; Indels 4; Gaps 2;
QY 361 TATACACTTAATTTTAAAGCTTTGAAATGGGCCGAGCCATAGCTCCTGCTT 420
DB 60529 TTTGAAATTTCCCTGATGAAAAAGTACAAAATGAGCCAGCCGTGGCTCAAGCCTGT 60588
QY 421 AATCCCAACACTTTGGAGGCCCAAGGTGGCGGATCACTGAGTCAAGATTTAAGACC 480
DB 60589 AATCCAGACACTTTGGAGGCCCAAGCGGGTGGCTCACTGAGTCAAGATTTAAGACC 60648
QY 481 AGCTTGGCAACATGGTGAACCTCTGCTTACTAAACGCAAAATTAAGCAAGGTGG 540
DB 60649 AGCTTGGCAACATGGTGAACCTCTGCTTACTAAACGCAAAATTAAGCAAGCATG 60708
QY 541 GTGGCAATGACCTGTGATCCCAACTCTCAGAGGTTGAGGAGGAATTCCTTAACC 600
DB 60709 GTGGCAATGACCTGTGATCCCAACTCTCAGAGGTTGAGGAGGAATTCCTTAACC 60768
QY 601 TAGAGGTGAGGTTGACAGTAAACCCAGAT--GTCACTGCACTCCAGCTT--GGCAACAG 656
DB 60769 TGGAGGCAAGAGTTGATTCATGATCAAGATCAAGCCATGCACTGCAAGCTTGGTATAG 60828
QY 657 AGCAAGCTTCATTAAGACACAAAGCTTTGAAATTTGTAATGATTTACTT 712
DB 60829 AGCAAGCTTCATTCATAAAAAAAATTAACAAAGTGGCGCGTTGTGTAGCT 60884
```

```
RESULT 41
ADQ97960.0/c
WP Sequence split into 4 fragments LOCUS ADQ97960 Accession Adq97960
WP Fragment Name Begin End
WP ADQ97960_0 1 110000
WP ADQ97960_1 100001 210000
WP ADQ97960_2 200001 310000
WP ADQ97960_3 300001 390183
ID ADQ97960 standard; DNA; 390183 BP.
AC ADQ97960;
XX 07-OCT-2004 (first entry)
DT
DE Human cancer associated sequence HD11-029, SEQ ID 937.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
OS Homo sapiens.
XX
XX WO2004060304-A2.
XX
XX 22-JUL-2004.
XX
XX 22-DEC-2003; 2003WO-US041389.
XX
XX 27-DEC-2002; 2002US-00330773.
XX
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX
XX WPI; 2004-543781/52.
XX
XX New isolated cancer associated nucleic acids comprising at least 10
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX cancers such as leukemia and lymphoma.
XX
XX Claim 1; SEQ ID NO 937; 199pp; English.
XX
XX The present invention relates to cancer associated sequences (ADQ97025-
```

CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 390183 BP; 107624A; 72906C; 76848G; 123530T; 0U; 92750Other;

Query Match 10.4%; Score 208; DB 12; Length 110000;  
Best Local Similarity 72.3%; Pred. No. 1.3e-33;  
Matches 298; Conservative 0; Mismatches 110; Indels 4; Gaps 2;

QY 275 TATGAATTAAGATTTCACAACTAGTACTGCTATTTGAAGATTAAAGCTTTGAAAG 394  
DB 64813 TTTAAATATAATACATGATATATTTTCATTTCTCAATTTTATGAGTGCAAAATTAAATTTTA 64754

QY 335 CAACATTTTAAATTAATTTTAAACTAATACCTTAATTTATTAAGCTTTGAAAG 394  
DB 64753 GAAATATGCAACTGATATTTGAACCTAATATATGACCTAAATTTATCAAAAATACACCT 64694

QY 395 GGCCGAGGCAAGTACTGCTGCTATATCCCAACCTTTGGGAGGCAAGGTGGCGGA 454  
DB 64693 AGCCGGCAAGGTGCTACGCTTATATCCCACTTTGGGAGGCAAGGTGGCGGA 64634

QY 455 TCACCTGAGGTGAGATTAAAGACCAAGCTGGCCAAATGCTGAAACCTGCTTACT 514  
DB 64633 TCACATGAGGTGAGATTAAAGACCAAGCTGGCCAAATGCTGAAACCTGCTTACT 64574

QY 515 AAAACGCAAAATTAGCCAGGTGTGTGCGATGACCTGTATGCCAATCTACAGAG 574  
DB 64573 AAAAATACAAAAATTCAGCCAGGCGTGTGTGCAACCTGTATGCCAATCTACAGAG 64514

QY 575 GTTGAAGGAGGAGAAATGCTTGAACCTAGAGAGGTGAGGTGAGTAACCCGAG--ATG 631  
DB 64513 GCTAAGGAGGAGAAATTCATTTGAACCTAGAGAGGTGAGGTGAGTAACCCGAG--ATG 64454

QY 632 TCACCTGACACTCCAGCTT-GGCAACAGAGCAAGATCCATTAAGACAAACAAA 682  
DB 64453 CCATTGCACTCCAGCTTGGGTGAGAGGCAAGATCCGCTTCAAAAAAAA 64402

RESULT 42  
ACN44154/c  
ID ACN44154 standard; DNA; 101209 BP.

XX ACN44154;  
XX  
XX 18-NOV-2004 (first entry)

DE Human genomic sequence hCG28567.  
XX  
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.

XX Homo sapiens.  
XX  
XX WO2003073826-A2.

XX 12-SEP-2003.  
XX  
XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.  
XX  
XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;  
XX  
XX WPI; 2003-328604/31.

XX  
XX PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX comprises a nucleotide sequence.  
XX  
XX Claim 1; SEQ ID NO 460; Opp; English.

XX The present invention relates to novel DNA and protein sequences which  
CC are associated with carcinomas. The sequences are useful for: (i) for  
CC screening drug candidates; (ii) for screening of bioactive agent capable  
CC of binding to Carcino Associated Protein (CAP); (iii) for screening of  
CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a bloodp;  
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
CC determining Carcino Associated (CA) gene copy number. In addition, the  
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
CC carcinoma including lymphoma. The present sequence is one such CA coding  
CC sequence. Note: This patent is an equivalent to basic patent  
CC US2002182586A1, for which no sequence data was published

XX  
SQ Sequence 101209 BP; 21450 A; 25161 C; 26356 G; 23759 T; 0 U; 4483 Other;

Query Match 10.3%; Score 207; DB 11; Length 101209;  
Best Local Similarity 59.5%; Pred. No. 2e-33;  
Matches 405; Conservative 0; Mismatches 270; Indels 6; Gaps 3;

QY 6 CACCGAGCTTGGGCTCCCAAGGTGGGATTACAGGCGGTGAGCCACAGCGCTGGTCA 65  
DB 25000 CACCGAGCTTGGGCTCCCAAGGTGGGATTACAGGCGGTGAGCCACAGCGCGCA 24941

QY 66 ATGTCTTATTAATTTGAAGAGCAAAATGGGCTTAAATCTGTCTTATTTGACAGCT 125  
DB 24940 ATGTGCTTCTTCCAGTACATGATGATGATTTGACAGCTTCTCAATCTGCTCAAGGCTT 24881

QY 126 TTGATGAGTCAATCCCAATGCTGCACTTACGAGCGCTTAAATGATTAATGCTCT 185  
DB 24880 GACAGCTCAATTCCTTTTCCATTAACAAATATTCCTTTCAGAGGAGCGCACCGTTG 24821

QY 186 CTCAGCTGCTTTGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 245  
DB 24820 CACATCACCCACACACAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 24761

QY 246 CTATGAAAAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 305  
DB 24760 TGGTTACAAAACCTGCGTGCAGGTTTCAATGGAAGATGTTTGAACATTTTAAATG 24701

QY 306 GCTATTGAAGATTAAAGTATTTATTAACAATTTTAAATTTTAAATTTTAAATTT 365  
DB 24700 TTGAAAAAATCCAGAGAGAAATTTTCAACACATGAAATTTATGAAATTTCACTT 24641

QY 366 CACTTAATTTAAAGCTTTGAATGGGCGGCGAGTACCTGCTGCTATCC 425  
DB 24640 TCAGT--ATCTTAATTAACGTAATGGGCTGGGCGGCTGCTGCTGCTATCC 24583

QY 426 CAACACTTGGGAGGCGCAAGGTGGGCGATCACTGAGGTCAAGAGTTTAAAGCAAGCT 485  
DB 24582 CAGCACTTGGGAGGCGGAGCGGCGGAGTTTACTGAGGTCAAGAAATTTGAAGCAAGCC 24523

QY 486 GGCCAACATGTAAGAACTCTGCTTCACTTAAAGAAAGCAAAATTTAGCAGGTGTGTC 545  
DB 24522 GGCCAACATGTAAGAACTCTGCTTCACTTAAAGAAAGCAAAATTTAGCAGGTGTGTC 24463

QY 546 ATGCACTGTAAGTCCCACTACTCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 605  
DB 24462 GGGCACTGTAAGTCCCACTACTCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 24403

QY 606 GGTGAGGTTGAGTAACCGAGAT--GTCACTGCACTCAGAGCTT-GGCAACAGAGCA 661  
DB 24402 GGTGAGGTTGAGTAACCGAGAT--GTCACTGCACTCAGAGCTT-GGCAACAGAGCA 24343

QY 662 GACTCCATAAAGCAACAAA 682  
DB 24342 GACTGCTTCAAAAAAAA 24322

RESULT 43  
ABD33366









This Page Blank (uspto)





Db 40680 TTAAACCAATACAAATAGTACTAGTCTCTGCTCCTTTTATTAGCACTAAATTTTGATTC 40739  
QY 1741 AATATTAATCTGTCATTTGAAGAGTTCCCTAATGTACTGTTCTAATCTTTCCTT 1800  
Db 40740 AATTAATTAATCTGTCATTTGAAGAGTTCCCTAATGTACTGTTCTAATCTTTCCTT 40799  
QY 1801 TCAAGGGGCGAGTGTCCCGTACACATAGCTAAATGGAGCTTCTTCAACCTAACCTATACC 1860  
Db 40800 TCAAGGGGCGAGTGTCCCGTACACATAGCTAAATGGAGCTTCTTCAACCTAACCTATACC 40859  
QY 1861 CAGAGGGCAGAACCTTAATATGCTGTGATGATGATCTTCTGCTTTCACTCTCAGCAGAGT 1920  
Db 40860 CAGAGGGCAGAACCTTAATATGCTGTGATGATGATCTTCTGCTTTCACTCTCAGCAGAGT 40919  
QY 1921 GTTGCAATTTGAGCTTCTGCGAGGGCCACCCGAGACCTTATCTGCTCAGAGTTTAACTCA 1980  
Db 40920 GTTGCAATTTGAGCTTCTGCGAGGGCCACCCGAGACCTTATCTGCTCAGAGTTTAACTCA 40979  
QY 1981 TCTAATTCAGTGAACACTTCA 2001  
Db 40980 TCTAATTCAGTGAACACTTCA 41000

## RESULT 2

US-09-078-294-4  
; Sequence 4, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 80246  
; TYPE: DNA  
; ORGANISM: Nucleotide sequence of NC-contig  
US-09-078-294-4

Query Match 97.9%; Score 1959.4; DB 3; Length 80246;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1988; Conservative 0; Mismatches 11; Indels 2; Gaps 2;  
QY 1 TGAATCCACGAGCTTGGAGCTCCCAAGTGTGGGATTTACAGAGCGTACACAGCGCTG 60  
Db 38738 TGAATCCACGAGCTTGGAGCTCCCAAGTGTGGGATTTACAGAGCGTACACAGCGCTG 38797  
QY 61 GTCGAATGCTTTATTTATTTGAAGAGCAACATGGGCTTTAATCTGTCTTCTATTGAC 120  
Db 38798 GTCGAATGCTTTATTTATTTGAAGAGCAACATGGGCTTTAATCTGTCTTCTATTGAC 38857  
QY 121 AGACTTTGATGAGTCAATCCCAATGCTGCCCTTACTGAAGCGCTTAAATGACTTGA 180  
Db 38858 AGACTTTGATGAGTCAATCCCAATGCTGCCCTTACTGAAGCGCTTAAATGACTTGA 38917  
QY 181 TCTCTCTCAGCTGTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 240  
Db 38918 TCTCTCTCAGCTGTCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 38976  
QY 241 TAAACCTATGAAAAGTGTGAGAGTGTGTTGATATGAAAATAGGATTTCAACAATG 300  
Db 38977 TAAACCTATGAAAAGTGTGAGAGTGTGTTGATATGAAAATAGGATTTCAACAATG 39035  
QY 301 TAGCTGCTATGAAAGTTTAAAGTATTTATTTAACAATTTTAAATTTTAAATTTTAAAC 360  
Db 39036 TAGCTGCTATGAAAGTTTAAAGTATTTATTTAACAATTTTAAATTTTAAATTTTAAAC 39095  
QY 361 TATATCACTTAATTTTAAAGAGCTTTGAATGGGCGAGGCGAGAGCTCTGCTGT 420  
Db 40739 TATATCACTTAATTTTAAAGAGCTTTGAATGGGCGAGGCGAGAGCTCTGCTGT 40799

Db 39096 TATATCACTTAATTTTAAAGAGCTTTGAATGGGCGAGGCGAGAGCTCTGCTGT 39155  
QY 421 AATCCCAACATTTGGAGCGCCAGGTGGCCGATCACTGAGTACAGAGTTTAAAGAC 480  
Db 39156 AATCCCAACATTTGGAGCGCCAGGTGGCCGATCACTGAGTACAGAGTTTAAAGAC 39215  
QY 481 AGCTGGCCAAACATGTGTAACCCCTGTCTTACTTAAAAAACGAAAATTTAGCAGGTGTG 540  
Db 39216 AGCTGGCCAAACATGTGTAACCCCTGTCTTACTTAAAAAACGAAAATTTAGCAGGTGTG 39275  
QY 541 GTGGCATGACCTGTGATGCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 600  
Db 39276 GTGGCATGACCTGTGATGCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 39335  
QY 601 TAGAGGTGAGAGTTGACATTAACCCGAGATGTACATGCACTCAGGCTGGCAACAGAGCA 660  
Db 39336 TAGAGGTGAGAGTTGACATTAACCCGAGATGTACATGCACTCAGGCTGGCAACAGAGCA 39395  
QY 661 AGACTCCATTAAGACAAACAAAGCTTTGAAATGTGTAATGAGTTGTACCTATCTTCA 720  
Db 39396 AGACTCCATTAAGACAAACAAAGCTTTGAAATGTGTAATGAGTTGTACCTATCTTCA 39455  
QY 721 TTAAAGAAATTCATCTTGTTCATTTTATTTTACTGACATGAGAGCTTCCAGCAATTTT 780  
Db 39456 TTAAAGAAATTCATCTTGTTCATTTTATTTTACTGACATGAGAGCTTCCAGCAATTTT 39515  
QY 781 AATTAAGCCCTCAGAGATTTATGTCACTGGCTATGTGATTAACAAATTTTGTCTTAA 840  
Db 39516 AATTAAGCCCTCAGAGATTTATGTCACTGGCTATGTGATTAACAAATTTTGTCTTAA 39575  
QY 841 TATATTTCTGCTCTTTTAAAGAAATGTCTCCCTAGAAACGTTGTGACCAACAT 900  
Db 39576 TATATTTCTGCTCTTTTAAAGAAATGTCTCCCTAGAAACGTTGTGACCAACAT 39635  
QY 901 ACACTGACTTTTACAAATAATCAGATGTGATGGCAACAGTTGAGATGTTTCAAAAGAT 960  
Db 39636 ACACTGACTTTTACAAATAATCAGATGTGATGGCAACAGTTGAGATGTTTCAAAAGAT 39695  
QY 961 TTTCAATTTGAAGGGGCCATTTGGGTTATTTAGATCTTGAAGCTGAATGCTTTGT 1020  
Db 39696 TTTCAATTTGAAGGGGCCATTTGGGTTATTTAGATCTTGAAGCTGAATGCTTTGT 39755  
QY 1021 TCTGTTTTCTGGCTTCTGGGAGAGAGAGAGATATTTCAAGCTTGTGATTT 1080  
Db 39756 TCTGTTTTCTGGCTTCTGGGAGAGAGAGATATTTCAAGCTTGTGATTT 39815  
QY 1081 TCTTTATCTTCATTTCAATACAGAGATGCTTCATATGACAGTGTGTCAAGTCAAT 1140  
Db 39816 TCTTTATCTTCATTTCAATACAGAGATGCTTCATATGACAGTGTGTCAAGTCAAT 39875  
QY 1141 CAAAAGAAAGAGAAACAGTTTCTTGTGTTTAAATTTTCAACCGGAAAGAAAGCAACCA 1200  
Db 39876 CAAAAGAAAGAGAAACAGTTTCTTGTGTTTAAATTTTCAACCGGAAAGAAAGCAACCA 39935  
QY 1201 TTTTGTCCGCTCTAATTTGAGAGTCAATGATGATGATGATGATGATGATGATGATG 1260  
Db 39936 TTTTGTCCGCTCTAATTTGAGAGTCAATGATGATGATGATGATGATGATGATGATG 39995  
QY 1261 GTGGTAAACAGGCTTCTTCAATATCTCAGCAGAGACTTGCATCTTCAATCTCTAGGC 1320  
Db 39996 GTGGTAAACAGGCTTCTTCAATATCTCAGCAGAGACTTGCATCTTCAATCTCTAGGC 40055  
QY 1321 TGAAGAAATGCTCAGAGAGATGAACAAATCTCAGAGCCCTTAACTGAAGCCAG 1380  
Db 40056 TGAAGAAATGCTCAGAGAGATGAACAAATCTCAGAGCCCTTAACTGAAGCCAG 40115  
QY 1381 TGTATTAAGCAACAGTCAAGAGGGTGAAGAACTAAAGTTTGAAGTCTCCCACTTCT 1440  
Db 40116 TGTATTAAGCAACAGTCAAGAGGGTGAAGAACTAAAGTTTGAAGTCTCCCACTTCT 40175  
QY 1441 TCTAGCTCAGAGAGCAAGTGAATTTTATTTAGTGAATTTTAAATTTTAAATTTA 1500  
Db 40176 TCTAGCTCAGAGAGCAAGTGAATTTTATTTAGTGAATTTTAAATTTTAAATTTA 40235



QY	1501	TTCTAAAGTCATGAAACAAGCCTATTATTAAGATAGTGGTCGTGAAGGTCGTAATAAC	1560
Db	40236	TTCTAAAGTCATGAAACAAGCCTATTATTAAGATAGTGGTCGTGAAGGTCGTAATAAC	40295
QY	1561	TCGATTTTACCAACCCCTCTTCTGTGAGGAAGCCATATGGAATCCTGTACAATGTTCA	1620
Db	40296	TCGATTTTACCAACCCCTCTTCTGTGAGGAAGCCAGATGGAATCCTGTAGATGTTCA	40355
QY	1621	TCCTACCAACGAACCTCTGTTTTCTATGTAGAGAAACAGAGGCCACAGTATTAACATATC	1680
Db	40356	TCCTACCAACGAACCTCTGTTTTCTATGTAGAGAAACAGAGGCCACAGTATGTAACATATC	40415
QY	1681	TTAAACCAATACAAATAGACTAGTGTCTGTGTCCTTTTATTAAGCACTAAATTTTGATCC	1740
Db	40416	TTAAACCAAGCAAAATAGACTAGTGTCTGTGTCCTTTTATTAAGCACTAAATTTTGATCC	40475
QY	1741	AATAATAAATGTGCCATTTGAAAGGACCTTCCCTAATATGATCTGGTCTAACTGTTCCCT	1800
Db	40476	AATAATAAATGTGCCATTTGAAAGGAGTTTCCCTAATATGATCTGGTCTAACTGTTCCCT	40535
QY	1801	TCAAGGGGCGCAGTGTCCCGTACACATAGCTAAATGGGACTTCTTCAACTACATTACC	1860
Db	40536	TCAAGGGGCGCAGTGTCCCGTACACATAGCTAAATGGGACTTCTTCAACTACATTACC	40595
QY	1861	CAGAGGGGAGAACCTTAATATGCTGTGAATACATTCGTGCTGTCACATTCGAGCAGACT	1920
Db	40596	CAGAGGGGAGAACCTTAATATGCTGTGAATGACATTCGTGCTGTCACATTCGAGCAGACT	40655
QY	1921	GTTGCATTGAGCTTGTGACAGGGCCACCCAGAGACTATATCTGCTCAGATGTTTAACTCA	1980
Db	40656	GTTGCATTGAGCTTGTGACAGGGCCACCCAGAGACTATATCTGCTCAGATGTTTAACTCA	40715
QY	1981	TCTAATTCAGTGAACACTTCA 2001	
Db	40716	TCTAATTCAGTGAACACTTCA 40736	

```

RESULT 3
US-09-949-016-14573/C
/ Sequence 14573, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14573
/ LENGTH: 393753
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)..(393753)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14573

```

Query Match	11.2%	Score 223.8	DB 3	Length 393753
Best Local Similarity	61.3%	Pred. No. 3.4e-45		
Matches 454	Conservative 0	Mismatches 267	Indels 20	Gaps 5

QY 1 TGAATCCACAGCCTTGCCCTCCCAAGTCTGGGATTAGAGCGTGAGCCACAGCGCTG 60

|||||

Db	122074	IGATTCACCCGCCCTTGGCTCCCAAAAGTGTGAATTCACAGAGTGAACCAACGCGCG	122015
Oy	61	GTGCAAGTCTTTATTTATTT--GAAGAGACAACATGGGCTTTAAATCTGCTCTTTATTTGA	119
Db	122014	GCCATATCATCTGTCTTCTTTTCAAGAAATGTAAGCTCTCTTCAAGGGCAGGGGTTTTGTCTT	121955
Oy	120	CAGACTTTGATGAGTGCMAATCCCAATGCTGCACCTTACTGACGGGCTTTAAATGACTTAA	179
Db	121954	GTTTGCTGCTTTATTTATTCATAATCTTAAATATGGCGCTTGCAATTAACACCAACATATG	121895
Oy	180	GTCTCTCTCAGCTGTCTTTCTGCATATGTAAAGTGGAAATTAATGATGCTTTTCAAGAGA	239
Db	121894	ATATTAAATGTAATGACTTAAGATGAAATTTAAACTCTGATGCTAAAAAGCTTCAGAGTTTC	121835
Oy	240	ATAAACCTATGAAAAAGTGTGAGATGATGTTTGTATGAAATTAAGATTTCAACAACTA	299
Db	121834	TTAG-----GTGTAATCTGTGTTTAAACACCAAGTTGATGATCACAAATTA	121788
Oy	300	GTAGTGTATTTGAAGATTTAAAGTTATTTTATTAACAATAATTTAAATTTTAAAAA	359
Db	121787	TTTAGT--AAATTAAGCGCAAAAAGACATTAAGTTTACCTGACTAATCTCATTTTGA	121730
Oy	360	CTAATACACTTAAATTTATTTAAAGACTTTGAAATGGGCGAGGCGCAGTAGCTCTGCTG	419
Db	121729	AGACGAAAAAGTGAAGTCAAAATGTGTTAAATGAGCGCAGAGGCGGTGCTCAATGCTG	121670
Oy	420	TAATCCCAACCTTGGAGGCGCAAGTGGGCGATCACTTGAGTCAAGAGTTTAAAGC	479
Db	121669	TAATCCGACACTTTAGAGGCGCAAGTGGGCGATCACTTGAGTCAAGAGTTTAAAGC	121610
Oy	480	CAGCTGGCCAAACATGTGMAAACCTGTCTCTAATAAACCCAAAATTTACCGAGTGT	539
Db	121609	CAGCTTGACCAACATGGGCAAAATCCGCTCTCACTAAAAATCAAAAAATTTACCGAGCAT	121550
Oy	540	GSTGGCAACACCTGTAGTCCCACTACTCAGAGGTTGAGGGAGAGAAATGCTTTGAAC	599
Db	121549	GSTGGCAACACTTTTAATCCAGCTGCTTGGAAGGCTGAGGAGAGAAATGCTTTGAAC	121490
Oy	600	CTAGAGGTGAGAGTTTGACGTAAACCGAGAT--GTCACTGCACTCCAGCCT--GGCAACA	655
Db	121489	CCGTGAGGTGAGAGTTTGACGTAAACCGAGAT--GTCACTGCACTCCAGCCTGGGTGAAC	121430
Oy	656	GAGCAAGACTCCATTAAGACAAACAAAGCTTTGAAATTTGTGAATGAGTTGTAACCTATC	715
Db	121429	GAGTGAAGACTCCGCTCAAAAAACAAAAACAAAAACAAAAACAAATTTGAGACATT	121370
Oy	716	TTCAATTAAGAAATTCATCTT 736	
Db	121369	GGAAAGTCTGAATCTCAGCAT 121349	

```

RESULT 4
US-09-949-016-14574/C
/ Sequence 14574, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14574
/ LENGTH: 393753
/ TYPE: DNA
/

```



Db 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 6  
US-09-949-016-14547/c  
Sequence 14547, Application US/09949016

Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14547  
LENGTH: 818128  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(818128)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14547

Query Match 11.2%; Score 223.8; DB 3; Length 818128;  
Best Local Similarity 61.3%; Pred. No. 4,4e-45;  
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY 1 TGAATCCAGCAGCTTGCGCTCCCAAGTGTGGGATTACAGGCGTGAGCCACGCGCTG 60  
DB 546449 TGAATCCAGCAGCTTGCGCTCCCAAGTGTGGGATTACAGGCGTGAGCCACGCGCG 546390  
QY 61 GTCGAATGCTTTATTTATTT-GAAGAGACAACATGGGCTTAAATCTGTCTTATTTGA 119  
DB 546389 GCTTATCATCTGCTTCTTTACAAGATGTAAGCTCCTGAGGGGAGGGGTTTGTCTT 546330  
QY 120 CAGACTTTGATGAGTCAATCCCATGCTGCCACTTACGAGCGGCTTAAATGACTTA 179  
DB 546329 GTTTGCTGCTTTATTTATCAATACCTAAATGGCGCTGGCATATTAACAGCCACATMG 546270  
QY 180 GTCCTCTCAGCTGTCTTTTGCAATATGTAAGTGAATATGATGGCTTTCAAGAGGA 239  
DB 546269 ATATTAATTAATGACTTAAGTAATTTTAACTCTAGCTAAAGTCTTCAGGGTTTC 546210  
QY 240 ATAAACCTATGAAAGTGTGAGGATAGTGTGATTAAGTAATTAAGATTTCACAAGTA 299  
DB 546209 TTAG-----GTGTAATCTGTTTAAACACCAAGTGAAGATTCACAGATTA 546163  
QY 300 GTAGCTGCTATTTGAAGATTAAAGATTATTTATTAACAATTTTAAATTTTAAAA 359  
DB 546162 TTAGAT--AATTAAGCGAAAAAAGACATTAAAGTTTACCTAGCTCAATACCTCATTTTGA 546105  
QY 360 CTAAATACCTAAATTTTAAAGCTTTGAAATGGGCGAGGCGAGTACCTGCTGCTG 419  
DB 546104 AGAAGAAAAAGTAAAGTCAAAATGTTAATATAGGCGAGGCGGCTGCTGCTGCTG 546045  
QY 420 TAAATCCACACTTTGGAGGCGCAAGGTGGCGGATCACTGAGTCAAGAGTTTAAAGAC 479  
DB 546044 TAAATCCACACTTTAGAGGCGCAAGGTGGCGGATCACTGAGTCAAGAGTTTAAAGAC 545985  
QY 480 CAGCTTGCCCAACTGTGTAACCTGTCTCTACTAAAAACGCAAAATTTAGCAGGTGT 539  
DB 545984 CAGCTTGCCCAACTGTGTAACCTGTCTCTACTAAAAACGCAAAATTTAGCAGGTGT 545925

QY 540 GATGGCATGACCTGTGATCCCACTACTCAGAGGTTGAGGAGGAGAAATTTGCTTGAAC 599  
DB 545924 GATGGCATGACCTGTGATCCCACTACTCAGAGGTTGAGGAGGAGAAATTTGCTTGAAC 545865  
QY 600 CTAGAGGTTGAGGTTGAGTGAACCCGAGAT--GTCACTGACCTCCAGCTT-GGCAACA 655  
DB 545864 CCGTAGGTTGAGGTTGAGTGAACCCGAGAT--GTCACTGACCTCCAGCTT-GGCAACA 545805  
QY 656 GAGCAAGACTCCATTAAGACAAACAAGCTTTGAATTTGTAAATGAGTTGACTATTC 715  
DB 545804 GAGTAGACTCCGCTCTCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACATT 545745  
QY 716 TTCAATTAAGAAATTCATCTT 736  
DB 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 7  
US-09-949-016-14548/c  
Sequence 14548, Application US/09949016

Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14548  
LENGTH: 818128  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(818128)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14548

Query Match 11.2%; Score 223.8; DB 3; Length 818128;  
Best Local Similarity 61.3%; Pred. No. 4,4e-45;  
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY 1 TGAATCCAGCAGCTTGCGCTCCCAAGTGTGGGATTACAGGCGTGAGCCACGCGCTG 60  
DB 546449 TGAATCCAGCAGCTTGCGCTCCCAAGTGTGGGATTACAGGCGTGAGCCACGCGCG 546390  
QY 61 GTCGAATGCTTTATTTATTT-GAAGAGACAACATGGGCTTAAATCTGTCTTATTTGA 119  
DB 546389 GCTTATCATCTGCTTCTTTACAAGATGTAAGCTCCTGAGGGGAGGGGTTTGTCTT 546330  
QY 120 CAGACTTTGATGAGTCAATCCCATGCTGCCACTTACGAGCGGCTTAAATGACTTA 179  
DB 546329 GTTTGCTGCTTTATTTATCAATACCTAAATGGCGCTGGCATATTAACAGCCACATMG 546270  
QY 180 GTCCTCTCAGCTGTCTTTTGCAATATGTAAGTGAATATGATGGCTTTCAAGAGGA 239  
DB 546269 ATATTAATTAATGACTTAAGTAATTTTAACTCTAGCTAAAGTCTTCAGGGTTTC 546210  
QY 240 ATAAACCTATGAAAGTGTGAGGATAGTGTGATTAAGTAATTAAGATTTCACAAGTA 299  
DB 546209 TTAG-----GTGTAATCTGTTTAAACACCAAGTGAAGATTCACAGATTA 546163  
QY 300 GTAGCTGCTATTTGAAGATTAAAGATTATTTATTAACAATTTTAAATTTTAAAA 359  
DB 546162 TTAGAT--AATTAAGCGAAAAAAGACATTAAAGTTTACCTAGCTCAATACCTCATTTTGA 546105

OY	360	CTAATACACTTAAATTATTATTAAGAGCTTTGAAATGGGCCAGGGCGAGTGCCTCGTCG	419
Db	546104	AGACGAAAAAAGTGAAGAGTCCTAAATATGTTAAATGAGGCCAGAGCGGTGCTATGCTCG	546045
OY	420	TAATCCCAACACTTTGGGAGGCCAAGTGGGCGGATCACCTGAGTCAGAGTTTAAGC	479
Db	546044	TAATCCAGCACTTTAGAGAGGCCAAGTGGGCGGATCACCTGAGTCAGGAGTTTGAGC	545985
OY	480	CAGCTGGGCCAACATGCTGAAACCTGTCTCTACTTAAAAACGCAAAATTTAGCCAGGTGT	539
Db	545984	CAGCTGGACCAACATGGCAAAATCCCGTCTTACTAAAAATACAAAATTTAGCCAGGCAT	545925
OY	540	GGTGGCATGCACCTGTAGTCCCACTACAGGAGTTGAGGGAGGAGAAATGCTTGAAC	599
Db	545924	GGTGGCAGGCACTTGTATATCCAGCTGCTTGGAGGCTGAGGAGGAGAAATGCTTGAAC	545865
OY	600	CTTAGAGGTGAGGTTTGACGTAATCCGAGAT---GTCACTGCACCTCAGGCT--GGCAACA	655
Db	545864	CGGTAGGTTGAGGTTTGACGTAAGCCAAATACGCGCATCTCAGGCTTGGTGACA	545805
OY	656	GAGCAAGACTCCATTAAGAACAACAAAGCTTTGAAATTGTGTAATGAGTTTACTATTC	715
Db	545804	GAGTGAAGACTCCGTCTCAAAAAACAAACAAACAAACAAACAAATTTGAGACCATTT	545745
OY	716	TTCAATTAAGAAATTCATCTT	736
Db	545744	GGAGATCTGAATCTCAGCAT	545724

RESULT 8  
US-09-94

```

;Sequence 14549, Application US/09949016
;Patent No. 6812339
;GENERAL INFORMATION:
;APPLICANT: VENTER, J. Craig et al.
;TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;FILE REFERENCE: CL001307
;CURRENT APPLICATION NUMBER: US/09/949, 016
;CURRENT FILING DATE: 2000-04-14
;PRIOR APPLICATION NUMBER: 60/241,755
;PRIOR FILING DATE: 2000-10-20
;PRIOR APPLICATION NUMBER: 60/237,768
;PRIOR FILING DATE: 2000-10-03
;PRIOR APPLICATION NUMBER: 60/231,498
;PRIOR FILING DATE: 2000-03-08
;NUMBER OF SEQ ID NOS: 207012
;SOFTWARE: FASTSEQ for Windows Version 4.0
;SEQ ID NO 14549
;LENGTH: 818128
;TYPE: DNA
;ORGANISM: Human
;FEATURE:
;NAME/KEY: misc feature
;LOCATION: (1) _(818128)
;OTHER INFORMATION: n = A,T,C or G
;S-09-949-016-14549

```

Query Match	11.2%;	Score 223.8;	DB 3;	Length 818128;
Best Local Similarity	61.3%;	Pred. No. 4.4e-45;		
Matches 454;	Conservative	0;	Mismatches 267;	Indels 20;
			Gaps	5

QY 1 TGATCCACGAGCTTGGGCTCCCAAGTCTGGGAATTA CAGGGGTATGGCACCAAGGCGCTG 60  
Db 546449 TGATCACCCCGCTTGGCTCCCAAGTCTGGGAATTA CAGGGGTATGGCACCAAGGCGCTG 546390  
QY 61 GTCGAATGTCCTTAATTAATT -GAAGAGACAACATGGGCGCTTAATCTGTCCTTATATTGA 119  
Db 546389 GCGTATCATCTCTCTTTCATTA CAAAGATGTAAAGCTCTCTGAGGGGAGGGGTTTGTCTT 546330  
QY 120 CAGACTTTGATGAGAGTCAAAATCCCAATGCTGCACCTTACTGAAAGGCTTAAATGACTTA 179

Db	546329	GTTTSCCTCTTATATCAATCTCTAAATGCGCGCTGGCATATTAACAGACCCACAATAG	546270
QY	180	GTCTCTCAGCTGTCTTTTTCGCAATATGTAAGGTGGAAATTAATGATGCTTTCAAGAGGA	239
Db	546269	ATATTAAATGGAATGACTAGAGTAATATTTTAAATCTCTAGCTAAAAAGTCTCAGGGTTTC	546210
QY	240	ATAAACCTATGAAAAGTGTGGAGATAGTGTGATATGTAATTAAGAAATTTTCAACAAGTA	299
Db	546209	TTAG-----GTGTAATCTGTTTAAACAACAAGTTGAGAAATCACAAGATT	546163
QY	300	GTAGCTGTATTGAAGATTTTAAGATTATTTATTAACAATATTATTAATAATTTTAAAAA	359
Db	546162	TTAGAT--AATTTAAGCGCAAAAAAGACATTAAATTACTAGCTCAATCTCATTTTGA	546105
QY	360	CTAATACCTTAAATATTTAAAGACTTTGAAATGGCGCAGCGCAGTATCTCTGCTTG	419
Db	546104	AGACAGAAAAAGTGAAGTCAAAATGGTTAAATGAGGCCAGAGGCGGTGCTCATAGCCTG	546045
QY	420	TAATCCCAACCTTTGGAGGCGCAAGTGGCGGATCACTAGGTCAAGAGTTTAAAG	479
Db	546044	TAATCCCAACCTTTAGAGAGCGCAAGTGGCGGATCACTAGGTCAAGAGTTTAAAG	545985
QY	480	CAGCTGCGCAACATGTTGAAACCTGTCTCTACTTAAACGCAAAATTTAGCCAGGTGT	539
Db	545984	CAGCTGCGCAACATGCGCAAAATCCCGTCTCTACTTAAAAATCAAAAATTTAGCCAGGCA	545925
QY	540	GGTGGCAAGCACTGTACTGCCCACTACTCAAGAGAGTTGAGGAGAGAAATTCCTTGAAC	599
Db	545924	GGTGGCAAGCACTGTATAATCCAGTGTCTTGGAGGCTGAGGCAAGAGAAATTCCTTGAAC	545865
QY	600	CTAGAGAGTGAAGTTTGACATTAACCCAGAT---GTCACTGACTTCAAGCTT-GGCACAA	655
Db	545864	CCGTGAGGTGAGGTTTGACAGTGAAGCAAGATTAACCCACTGCACTCCAGCTGGGGTGA	545805
QY	656	GAGCAGACCTCCATTAAGACAAACAAGCTTTGAAATTTGTAAATGAGTTTATCTATC	715
Db	545804	GAGTCAGACTCCGCTCTCAAAAAACAACAACAACAACAACAACAATTTGAGACATTT	545745
QY	716	TTCAATTTAAGAAATTCATCTT 736	
Db	545744	GGAGATCTGAATCTCAGCAT 545724	

## RESULT 5

```

US-09-949-016-14550/c
; Sequence 14550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14550
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14550

```

Query Match	11.2%;	Score 223.8;	DB 3;	Length 818128;
-------------	--------	--------------	-------	----------------

Best Local Similarity 61.3%; Pred. No. 4.4e-45;  
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY 1 TGATCCACGAGCTTGGCTCCCAAGTCTGGAGTATTAAGCGCTGAGCACCAGCCCTG 60  
Db 546449 TGATCCACGAGCTTGGCTCCCAAGTCTGGAGTATTAAGCGCTGAGCACCAGCCCG 546390

QY 61 GTCGATGCTTATTAATTT-GAAGAGCAACATGGGCTTAAATCTGCTTATTTGA 119  
Db 546389 GCTTATCATCTGCTCTTTTACAGAGATGTAAGCTCTCTGAGGGGAGGGTTTGTCTT 546330

QY 120 CAGACTTGTATGAGTCAATCCCAATGCTGCCACTTACTGAACGCGCTTAAATGACTTA 179  
Db 546329 GTTTCGCTTATTAATCAATACCTAAATGGCGCTGGCATATTAACAGCCACAATAG 546270

QY 180 GTCTCTCAGCTGCTTTCTGATATGTAAGTGAATATATGATGCTTTCAAGAGA 239  
Db 546269 ATATTAAATGAATGACTAAAGTAATTTAACTCTAGCTAAAGTCTTCAAGGTTTC 546210

QY 240 ATAAACCTATGAAAAGTGTGAGATGTTGATGATGATGAATTAAGATTTCAACAAGTA 299  
Db 546209 TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAGATCAAGATTA 546163

QY 300 GTAGCTGCTATGAGATTTAAGATTTATTAACAATTTAAATTTAAATTTAAAA 359  
Db 546162 TTAGAT--AATTAAACGAAAAAGACATTAAGTTTAACTAGCTCAATACCTCATTTTGA 546105

QY 360 CTAAATCACTTAATTTATTAAGAGCTTTGAATGGGCCAGGCGAGTGTCTGCGCTG 419  
Db 546104 AGACAGAAAAAGTAAAAAGTCAAAATGTTAAATGAAGCCAGGCGGTGCTCATGCTG 546045

QY 420 TAAATCCCAACACTTTGGAGGCAAGAGTGGGCGGATCACTGAGTCAAGAGTTTAAAG 479  
Db 546044 TAAATCCCAACACTTTGGAGGCAAGAGTGGGCGGATCACTGAGTCAAGAGTTTAAAG 545985

QY 480 CAGCTGGCCCAACATGCTGTAACCTGTCTTAATAAAACGAAAAATTTAGCCAGGTG 539  
Db 545984 CAGCTGGCCCAACATGCTGTAACCTGTCTTAATAAAATTAAGCCAGGCAAT 545925

QY 540 GGTGGCATGCACTGTATGTCCTCACTAAGAGGTTGGAGAGAGAAATTCCTTGAAC 599  
Db 545924 GGTGGCATGCACTGTATGTCCTCACTAAGAGGTTGGAGAGAGAAATTCCTTGAAC 545865

QY 600 CTAGAGGTGGAGGTTGAGTAACCCGAGAT---GTCACTGCACTCCAGCT--GGCAACA 655  
Db 545864 CCGTAGGTGGAGGTTGAGTAACCCGAGAT---GTCACTGCACTCCAGCT--GGGAGACA 545805

QY 656 GAGCAAGACTCCATAAAGCAACAAAGCTTTGAAATTTGTAAATGATGTTTACTATC 715  
Db 545804 GAGTAGAGCTCCGTCAAAAAACAAACAAACAAACAAACAAATTTGAGACCAT 545745

QY 716 TTCAATTTAAGAAATTCATCTT 736  
Db 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 10  
US-09-949-016-14551/c  
; Sequence 14551, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTUR, J. Craig et al.  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14551  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1) ... (818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14551

Query Match 11.2%; Score 223.8; DB 3; Length 818128;  
Best Local Similarity 61.3%; Pred. No. 4.4e-45;  
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY 1 TGATCCACGAGCTTGGCTCCCAAGTCTGGAGTATTAAGCGCTGAGCACCAGCCCTG 60  
Db 546449 TGATCCACGAGCTTGGCTCCCAAGTCTGGAGTATTAAGCGCTGAGCACCAGCCCG 546390

QY 61 GTCGATGCTTATTAATTT-GAAGAGCAACATGGGCTTAAATCTGCTTATTTGA 119  
Db 546389 GCTTATCATCTGCTCTTTTACAGAGATGTAAGCTCTCTGAGGGGAGGGTTTGTCTT 546330

QY 120 CAGACTTGTATGAGTCAATCCCAATGCTGCCACTTACTGAACGCGCTTAAATGACTTA 179  
Db 546329 GTTTCGCTTATTAATCAATACCTAAATGGCGCTGGCATATTAACAGCCACAATAG 546270

QY 180 GTCTCTCAGCTGCTTTCTGATATGTAAGTGAATATATGATGCTTTCAAGAGA 239  
Db 546269 ATATTAAATGAATGACTAAAGTAATTTAACTCTAGCTAAAGTCTTCAAGGTTTC 546210

QY 240 ATAAACCTATGAAAAGTGTGAGATGTTGATGATGATGAATTAAGATTTCAACAAGTA 299  
Db 546209 TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAGATCAAGATTA 546163

QY 300 GTAGCTGCTATGAGATTTAAGATTTATTAACAATTTAAATTTAAATTTAAAA 359  
Db 546162 TTAGAT--AATTAAACGAAAAAGACATTAAGTTTAACTAGCTCAATACCTCATTTTGA 546105

QY 360 CTAAATCACTTAATTTATTAAGAGCTTTGAATGGGCCAGGCGAGTGTCTGCGCTG 419  
Db 546104 AGACAGAAAAAGTAAAAAGTCAAAATGTTAAATGAAGCCAGGCGGTGCTCATGCTG 546045

QY 420 TAAATCCCAACACTTTGGAGGCAAGAGTGGGCGGATCACTGAGTCAAGAGTTTAAAG 479  
Db 546044 TAAATCCCAACACTTTGGAGGCAAGAGTGGGCGGATCACTGAGTCAAGAGTTTAAAG 545985

QY 480 CAGCTGGCCCAACATGCTGTAACCTGTCTTAATAAAACGAAAAATTTAGCCAGGTG 539  
Db 545984 CAGCTGGCCCAACATGCTGTAACCTGTCTTAATAAAATTAAGCCAGGCAAT 545925

QY 540 GGTGGCATGCACTGTATGTCCTCACTAAGAGGTTGGAGAGAGAAATTCCTTGAAC 599  
Db 545924 GGTGGCATGCACTGTATGTCCTCACTAAGAGGTTGGAGAGAGAAATTCCTTGAAC 545865

QY 600 CTAGAGGTGGAGGTTGAGTAACCCGAGAT---GTCACTGCACTCCAGCT--GGCAACA 655  
Db 545864 CCGTAGGTGGAGGTTGAGTAACCCGAGAT---GTCACTGCACTCCAGCT--GGGAGACA 545805

QY 656 GAGCAAGACTCCATAAAGCAACAAAGCTTTGAAATTTGTAAATGATGTTTACTATC 715  
Db 545804 GAGTAGAGCTCCGTCAAAAAACAAACAAACAAACAAACAAATTTGAGACCAT 545745

QY 716 TTCAATTTAAGAAATTCATCTT 736  
Db 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 11  
US-09-949-016-14552/c  
; Sequence 14552, Application US/09949016  
; Patent No. 6812339

```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14552
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552
```

```

Query Match      11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4.4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;
```

```

QY      1  TGATCCAGCAGCTTGCGCTGCCAAAGTCTGGAGTTAAGCGCGTGAAGCCGCTG 60
DB      546449  TGATCCAGCAGCTTGCGCTGCCAAAGTCTGGAGTTAAGCGCGTGAAGCCGCTG 546390
QY      61  GTGCAATGCTTTATTATT- GAAGACAACATGCGCTTAATCTGCTTATTTA 119
DB      546389  GCTTATCATCTGCTCTTACCAAGATGTAAGCTCTCGAGGCGAGGGGTTTGTCTT 546330
QY      120  CAGACTTGATGAGATCAATCCCAATGCTGCCACTTACTGAAGCGCTTAATGACTTA 179
DB      546329  GTTTGCTGCTTATTATCAATCAATGCTGAATGGCGCTGCATTAACAGCCCAATATG 546270
QY      180  GTCTCTCAGCTGCTTTCTGCATATGTAAGTGAATATGATGCTTTCAAGAGGA 239
DB      546269  AATTAATTTGATGACTAGTGAATTTAACTCTAGCTAAAGTCTTCAAGGTTTC 546210
QY      240  ATTAACCTATGAAAAGTGTGAGATAGTGTGATTAATAATAGATTTCAACAAGTA 299
DB      546209  TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAGATCAAGATTA 546163
QY      300  GTAGCTGCTATGAAGATTTAAGATTTATTATTAACATTTTAATAATTTTAAAA 359
DB      546162  TTAGAT--AATTAAGCGAAAAAGACATTAGTTTACTAGTCAATCTCATTTTGA 546105
QY      360  CTAAATCACTTAATATTATAAGAGCTTGAATGGCGCAGCGAGTCTGCTGCG 419
DB      546104  AACAAGAAAAAGTGAAGTCAAAATGTTAAATGAGGCCAGAGCGGTGCTCATGCTG 546045
QY      420  TAATCCCAACACTTTGGAGGCGCAAGTGGCGGATCACTTGAAGTCAAGAGTTTAAAG 479
DB      546044  TAATCCCAACACTTTAGAGGCGCAAGTGGCGGATCACTTGAAGTCAAGAGTTTAAAG 545985
QY      480  CAGCTTGCGCAACATGCTGAACCTGTCTCTAATTAAGCAAAATTTAGCCAGGTGT 539
DB      545984  CAGCTTGCAACATGCGCAAAATCCGCTCTTAATAAATCAAAATTTAGCCAGGCAT 545925
QY      540  GGTGCGATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGGAGAAATGCTTGA 599
DB      545924  GGTGCGAGGCACTTGTAAATCCAGCTGCTTGGAGGCTAGGCGAGGAGAAATGCTTGA 545865
QY      600  CTAGAGGTTGAGGTTGAGTAAATCCAGAT--GTCACTGCACTTCAAGCTT-GGCAACA 655
DB      545864  CCGTAGGTTGAGGTTGAGTAAATCCAGATCAAGATCAAGCTTCAAGCTTGGAGTGA 545805
```

```

QY      656  GAGCAAGCTCCATTAAGACAACAAAGCTTGAATTTGTAATGATGTTACTATC 715
DB      545804  GAGTGAGATCTCGCTCAAAAAACAAACAAACAAACAAACAAACAAACAAACAAAC 545745
QY      716  TTCATTTAAGAAATTCATCTT 736
DB      545744  GGAAGATCTGAATCTCAGCAT 545724
```

## RESULT 12

```

US-09-949-016-14553/c
; Sequence 14553, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14553
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14553
```

```

Query Match      11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4.4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;
```

```

QY      1  TGATCCAGCAGCTTGCGCTGCCAAAGTCTGGAGTTAAGCGCGTGAAGCCGCTG 60
DB      546449  TGATCCAGCAGCTTGCGCTGCCAAAGTCTGGAGTTAAGCGCGTGAAGCCGCTG 546390
QY      61  GTGCAATGCTTTATTATT- GAAGACAACATGCGCTTAATCTGCTTATTTTA 119
DB      546389  GCTTATCATCTGCTCTTACCAAGATGTAAGCTCTCGAGGCGAGGGGTTTGTCTT 546330
QY      120  CAGACTTGATGAGATCAATCCCAATGCTGCCACTTACTGAAGCGCTTAATGACTTA 179
DB      546329  GTTTGCTGCTTATTATCAATCAATGCTGAATGGCGCTGCATTAACAGCCCAATATG 546270
QY      180  GTCTCTCAGCTGCTTTCTGCATATGTAAGTGAATATGATGCTTTCAAGAGGA 239
DB      546269  AATTAATTTGATGACTAGTGAATTTAACTCTAGCTAAAGTCTTCAAGGTTTC 546210
QY      240  ATTAACCTATGAAAAGTGTGAGATAGTGTGATTAATAATAGATTTCAACAAGTA 299
DB      546209  TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAGATCAAGATTA 546163
QY      300  GTAGCTGCTATGAAGATTTAAGATTTATTATTAACATTTTAATAATTTTAAAA 359
DB      546162  TTAGAT--AATTAAGCGAAAAAGACATTAGTTTACTAGTCAATCTCATTTTGA 546105
QY      360  CTAAATCACTTAATATTATAAGAGCTTGAATGGCGCAGCGAGTCTGCTGCG 419
DB      546104  AACAAGAAAAAGTGAAGTCAAAATGTTAAATGAGGCCAGAGCGGTGCTCATGCTG 546045
QY      420  TAATCCCAACACTTTGGAGGCGCAAGTGGCGGATCACTTGAAGTCAAGAGTTTAAAG 479
DB      546044  TAATCCCAACACTTTAGAGGCGCAAGTGGCGGATCACTTGAAGTCAAGAGTTTAAAG 545985
```

QY	480	CAGCTGCGCCAA	CATGTGTAAC	CCCTGTCTCTA	CTAAAAACGCAAAATTTAGCCAGGTGT	539	
Db	545984	CAGCTTGACCA	CATGGCAAAAT	CCCGTCTCTA	CTAAAAATTCMAAAATTTAGCCAGGCAT	545925	
QY	540	GGTGCATGCA	CTGTAGTCCCA	CTA	CTCAGAGGTTGAGGAGGAGATTTGTTTGAAC	599	
Db	545924	GCTGGCAGGCA	CTTGTAA	TCCAGCTGCTT	TGGGAGGCTGAGGAGGAGAAATTTGTTGAAC	545865	
QY	600	CTAGAGGTGAG	GTTCGACGTAA	CCCGAGAT---	GTCAC	CTGCACTCCAGCCTTGGCAACA	655
Db	545864	CCGTAGGGTGAG	GTTCGACGTAG	CCCAATATCA	CCGCACTGTCA	CTTCACCTCTGGGAGACA	545805
QY	656	GAGCAAGACT	CCATTAAGA	CAACAAAG	CTTTGAATTTGTGTAATGAGTTGTACTATC	715	
Db	545804	CAGTGTAGACT	CCGCTCC	CAAAAAACAAAACAAAAC	CAAAACCAAAACCAAAACCAAAACATTTGAGACAT	545745	
QY	716	TTCA	TTTAAAGAA	ATTCATCTT	736		
Db	545744	GGAA	GATCTGAA	ATTCAGCAT	545724		

```

RESULT 13
US-09-949-016-14554/c
; Sequence 14554, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14554
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554

```

Query Match	11.28;	Score 223.8;	DB 3;	Length 818128;
Best Local Similarity	61.3%;	Pred. No. 4,4e-45;		
Matches 454;	Conservative 0;	Mismatches 267;	Indels 20;	Gaps 5;
QY	1	TCATCCACCACCTTGCCCTCCCAAGTGTGCGATTACAGCGCGTAGCCACACCGCTTG 60		
Db	546449	TGATCCACCCGCTTGCCCTCCCAAGTGTGCGATTACAGCGTAGCCACACCGCTTG 60		
QY	61	GTGCAATGCTTTATTATT- GAAGAGACAACATGGCCCTTAAATCTGCTCTTATTGA 119		
Db	546389	GCTATCATCTGCTCTTCTTTACAAAGATGTAACTGCTCTGAGGGCAGGGGTTTGCTT 119		
QY	120	CAGACTTTGATGAGATCAATCCCATGCTGCCACTTACTGAAAGCGCTTAAATGACTTA 179		
Db	546329	GTTTGTGCTTTATTATCAATACCTTAAATGGCGCCCTGGCATATTAACAGCCACATAG 179		
QY	180	GTCCTCTCAGCTGCTTTCTCGCATATGATAGGAAATATATATGCTTTCAAGAGGA 239		
Db	546269	ATATTATTAATGACTTAAGTGAATATTTAACTCTTAGCTAAAAAGCTTCAAGGTTTC 239		
QY	240	ATAAACCTATGAAAAAGTGTGAGGATAGTGTGATATGAAATTAAGGATTTCAACAGTA 299		

Db	546209	TTAG-----GTGAATCTGGTTTAAACAACCAAGTTGAGAAATCAAGAAATTA	546163
Qy	300	GTAGCTGATATGAGATTATTAAGATTATTTATTAACAATATTTAATAAATTTTAAAAA	359
Db	546162	TTAGAT--AAITTAAGCGAAAAAAGACAATTAAAGTTTACCTTAGCTCAATACTCATTTTGA	546105
Qy	360	CTAATATCACTTAAATATTAATAAGCCTTGAAATGCGGCGAGCGAGTATGCTCGTCCGTG	419
Db	546104	AGACAGAAAAAGTGAAAGTCAAAATATGTTAAATGAGGCCAGAGCGGTGGCTCAATCCCTG	546045
Qy	420	TAATCCCAACCTTTGGAGGCCAAGTGGCGGATCACTTGAGGTCAGAGTTTAAAGC	479
Db	546044	TAATCCGACACTTTAGAGGCGCAAGTGGCGGATCACTTGAGGTCAGAGTTTGAAGAC	545985
Qy	480	CAGCTTGGCCAAACATGTTGAATCCCTGTCTCACTAAAAAGCAAAAAATTTAGCAGGTGT	539
Db	545984	CAGCTTGACCAACATGCAAAATCCCGTCTCACTAAAAATCAAAAAATTTAGCAGGCAT	545925
Qy	540	GGTGCATGACACTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTTGAAC	599
Db	545924	GGTGCAGGCACTTGTAATCCAGCTGCTTGAGAGGCTGAGGCAAGAGAAATGCTTTGAAC	545865
Qy	600	CTAGAGGTGAGGTTTGCAGTAAACCGAGAT---GTCACTGCAATCCAGACCT--GGCAACA	655
Db	545864	CCGTAGGTGAGGTTTGCAGTAAACCGAGATCAACGCACTGGCACTCCAGCTGGGGTGAAC	545805
Qy	656	GAGCAAGACTCATATAAGACAACAAGCCTTGAAATGTTGTAATGAGTTTGAACCTATC	715
Db	545804	GAGTGAAGACTCCGCTCAAAAAACAACCAAAAAACAACCAAAAAAATTTGAGACCATT	545745
Qy	716	TTCAITTAAGAAATTCATCTT 736	
Db	545744	GGAAAGATCTGAATCTCAGCAT 545724	

```

RESULT 14
US-09-949-016-14555/c
; Sequence 14555, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14555
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14555

```

Query Match	11.2%	Score 223.8	DB 3	length 818128
Best Local Similarity	61.3%	Pred: 1.4e-45		
Matches 454	Conservative 0	Mismatches 267	Indels 20	Gaps 5
QY	1	TGATCCACCAAGCTTGGCTCTCCCAAGATGCTGGAGATTAAAGGCGGTAGACCAACAGCCCTG	60	
DB	546449	TGATCCACCGCTTGGCTCTCCCAAGATGCTGGAGATTAAAGGCGGTAGACCAACAGCCCTG	546390	
QY	61	GTGCAATGCTTTATTATTATTT-GAAGAGACCAACAGGCGCTTAAATCTGTCTTCTATTATTA	119	



```

; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14556
Query Match 11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4,4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

1 TGATCCACAGCCTTGCCCTCCCAAGTGTGGATTTACAGGGTGTAGCCACACCCCTG 60
546449 TGATCCACCGCTTGCCCTCCCAAGTGTGGATTTACAGGGTGTAGCCACACCCCG 546390
61 GTGCAATGCTTTATTAATTT-GAAGACAACAATGGCCCTTAATCTGCTTATTTGA 119
546389 GCTTATCATCTGCTTCTTTTACAGATGTAACTCTCTGAGGGAGGGGTTTGTCTT 546330
300 GTAGCTGCTATTGAAGATTAAAGATTATTATTACACTATTAAATTTTAAAAA 359
546162 TTAGAT--AATTAAAGCGAAAAAGACATTAAAGTTTAACTTAACTCAATTTTGA 546105
360 CTAAATCACTTAATTAATTAAAGAGCTTGAATGGGCCAGGCGAGTAGCTCTGCTG 419
546104 AGACAGAAAAAGTAAAGTCAAAATGTAAATGAGGCCAGAGCGGTGCTATGCTG 546045
420 TAATCCCAACACTTTGGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGC 479
546044 TAATCCAGCACTTTAGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGC 545985
480 CAGCTGGCCAACTGTGTAAACCTGTCTTACTTAAAAACGAAAAATTAGCCAGGTG 539
545984 CAGCTGGCCAACTGTGTAAACCTGTCTTACTTAAAAATTAGCCAGGTG 545925
540 GGTGGCATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGGAATTCCTGAAC 599
545924 GGTGGAGGCACTGTGAATCCAGCTGTGGAGGCTGAGGAGGAATTCCTGAAC 545865
600 CTAGAGGTGAGGTGTCAGTAACCCGAGAT--GTCACTGCACTCAGCTT-GGCAACA 655
545864 CCGTAGGTGAGGTGTCAGTAACCCGAGATCAAGCACTGCACTCAGCTTGGTGACA 545805
656 GAGCAAGACTCCATTAAGACAACAAGCTTTGAAATTTGTAAATGAGTTTACTTATC 715
545804 GAGTGAGACTCCGTCTCAAAAAACAACAACAACAACAACAATTGAGACATT 545745
716 TTCATTTAAGAAATTCATCTT 736
545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 15
US-09-949-016-14556/C
; Sequence 14556, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14556
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14556
Query Match 11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4,4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

1 TGATCCACAGCCTTGCCCTCCCAAGTGTGGATTTACAGGGTGTAGCCACACCCCTG 60
546449 TGATCCACCGCTTGCCCTCCCAAGTGTGGATTTACAGGGTGTAGCCACACCCCG 546390
61 GTGCAATGCTTTATTAATTT-GAAGACAACAATGGCCCTTAATCTGCTTATTTGA 119
546389 GCTTATCATCTGCTTCTTTTACAGATGTAACTCTCTGAGGGAGGGGTTTGTCTT 546330
120 CAGACTTGAATGAGTCAAAATCCCAATGTCGCCACTTACTGAACGCGCTTAAATGA 179
546329 GTTGGCTGCTTATTAATCAATCTAAATGGCGCTGGCATTTAACAGACCAATAG 546270
180 GTCTCTCAGCTGTCTTTCTGATATGTAAGTGAATATATGATGCTTTCAAGAGGA 239
546269 ATATTAAATGAATGACTTAAGTAAATTTAACTCTTAAATTAAGTCTTCAGGGTTC 546210
240 ATAAACCTAAGAAAGTGTGAGATAGTGTGATGATGAATTAAGATTTCAACAAGTA 299
546209 TTAG-----GTGTAATCTGTTTAAACATTAACACACCAAGTTGAAATCACAGATT 546163
300 GTAGCTGCTATTGAAGATTAAAGATTATTATTACACTATTAAATTTTAAAAA 359
546162 TTAGAT--AATTAAAGCGAAAAAGACATTAAAGTTTAACTTAACTCAATTTTGA 546105
360 CTAAATCACTTAATTAATTAAAGAGCTTGAATGGGCCAAGTGGCGGATCACTGAGTCA 419
546104 AGACAGAAAAAGTAAAGTCAAAATGTAAATGAGGCCAAGAGCGGTGCTATGCTG 546045
420 TAATCCCAACACTTTGGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGC 479
546044 TAATCCAGCACTTTAGAGGCCAAGTGGCGGATCACTGAGTCAAGAGTTTAAAGC 545985
480 CAGCTGGCCAACTGTGTAAACCTGTCTTACTTAAAAACGAAAAATTAGCCAGGTG 539
545984 CAGCTGGCCAACTGTGTAAACCTGTCTTACTTAAAAATTAGCCAGGTG 545925
540 GGTGGCATGCACTGTAGTCCCAACTACTCAGAGGTTGAGGAGGAATTCCTGAAC 599
545924 GGTGGAGGCACTGTGAATCCAGCTGTGGAGGCTGAGGAGGAATTCCTGAAC 545865
600 CTAGAGGTGAGGTGTCAGTAACCCGAGAT--GTCACTGCACTCAGCTT-GGCAACA 655
545864 CCGTAGGTGAGGTGTCAGTAACCCGAGATCAAGCACTGCACTCAGCTTGGTGACA 545805
656 GAGCAAGACTCCATTAAGACAACAAGCTTTGAAATTTGTAAATGAGTTTACTTATC 715
545804 GAGTGAGACTCCGTCTCAAAAAACAACAACAACAACAACAATTGAGACATT 545745
716 TTCATTTAAGAAATTCATCTT 736
545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 16
US-09-949-016-14557/C
; Sequence 14557, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14557
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14557

Query Match      11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4,4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY      1  TGATCCACGAGCTTGCGCTCCCAAGTGGGATTTACAGCGGTGAGCCACCGCTG 60
DB      546449  TGATCCACCGGCTTGCGCTCCCAAGTGGGATTTACAGAGTGAGCCACCGCCG 546390

QY      61  GTGCAATGCTTTATTTATTT-GAAGAGACAACATGGGCGCTTAAATCTGCTTATTTGA 119
DB      546389  GCGTATCATCTGCTCTTTTACCAAGATGTAAGCTCTCTGAGGCGAGGGGTTTGCTTT 546330

QY      120  CAGACTTGATGAGTCAATCCCAATGCTGCACCTTACTGAAACGCGCTTAAATGACTTA 179
DB      546329  GTTTGCTGCTTTATTTATTTCAATCTTAATGGCGCTGCGCATTTACAGCCACATAG 546270

QY      180  GTCTCTGAGCTGTCTTCTGCATATGTAAGTGAATATGATGCGCTTCAAGAGCA 239
DB      546269  ATATTATTTGAATGACTAAGTGAATATTTAACTCTAGCTAAAGTCTTTCAGGGTTTC 546210

QY      240  ATTAACCTATGAAAAGTGTGAGAGATAGCTTTGATATGTAATGAAATTTCAACAAGTA 299
DB      546209  TTAG-----GTGTAATCTGGTTTAAACACCCAGTTGGAATCACAGATTA 546163

QY      300  GTAGCTGTATTTGAAGATTTAAGATTTATTTATTTACACTATTTAAATTTTAAAAA 359
DB      546162  TTAGAT--AATTTAAGGAAAAAAGACATTTAAGTTTACCTAGCTCATATCTCATTTGA 546105

QY      360  CTAAATCACTTAATTTAATAAGACTTTGAAAATGGCCAGCGCGAGTACCTCTGCTG 419
DB      546104  AGACAGAAAAAGTGAAGTCAAAATGTTAATGAGGCCAGAGCGGCTCATGCTG 546045

QY      420  TAAATCCACACTTTGGGAGGCCCAAGTGGCGGATCACTTGAGTCAAGAGTTTAAAG 479
DB      546044  TAAATCCACAGACTTTGAGAGGCCCAAGTGGCGGATCACTTGAGTCAAGAGTTTGAAC 545985

QY      480  CAGCTGGCCCAAGATGGTGAACCCCTGCTCTACTTAAAAACGCAAAATTTAGCCAGGTG 539
DB      545984  CAGCTGGCCCAAGATGGCAAAATCCGCTCTCTACTTAAAAATCAAAATTTAGCCAGCAT 545925

QY      540  GGTGCAATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGGAAATTTGTTGAAC 599
DB      545924  GGTGGCAGGCACTTTGATCCAGCTGCTTGGGAGGCTGAGGCGAGGAAATTTGTTAAAC 545865

QY      600  CTAGAGGTGAGGTTTGCAATACCCGAGAT--GTCACTGCACTCCAGCTT-GGCAACA 655
DB      545864  CCGTAGGTGAGGTTTGCAAGTCAAGATCAACGCCACTGCACTCCAGCTTGGGTGACA 545805

QY      656  GAGGAAGACTCCATAAAGACAACAAGCTTTGAATTTGTATATGATGTTGTATCTATC 715
DB      545804  GAGGGAAGACTCCGCTTCAAAAAACAACAACAACAACAACAATTTTGAAGCAATT 545745

QY      716  TTCAATTAAGAAATTCATCTT 736
DB      545744  GGAAGATCTGATCTCAGCAT 545724
```

```

RESULT 17
US-09-949-016-14558/c
; Sequence 14558, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14558
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14558

Query Match      11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4,4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY      1  TGATCCACGAGCTTGCGCTCCCAAGTGGGATTTACAGCGGTGAGCCACCGCTG 60
DB      546449  TGATCCACCGGCTTGCGCTCCCAAGTGGGATTTACAGAGTGAGCCACCGCCG 546390

QY      61  GTGCAATGCTTTATTTATTT-GAAGAGACAACATGGGCGCTTAAATCTGCTTATTTGA 119
DB      546389  GCGTATCATCTGCTCTTTTACCAAGATGTAAGCTCTCTGAGGCGAGGGGTTTGCTTT 546330

QY      120  CAGACTTGATGAGTCAATCCCAATGCTGCACCTTACTGAAACGCGCTTAAATGACTTA 179
DB      546329  GTTTGCTGCTTTATTTATTTCAATCTTAATGGCGCTGCGCATTTACAGCCACATAG 546270

QY      180  GTCTCTGAGCTGTCTTCTGCATATGTAAGTGAATATGATGCGCTTCAAGAGCA 239
DB      546269  ATATTATTTGAATGACTAAGTGAATATTTAACTCTAGCTAAAGTCTTTCAGGGTTTC 546210

QY      240  ATTAACCTATGAAAAGTGTGAGAGATAGCTTTGATATGTAATGAAATTTCAACAAGTA 299
DB      546209  TTAG-----GTGTAATCTGGTTTAAACACCCAGTTGGAATCACAGATTA 546163

QY      300  GTAGCTGTATTTGAAGATTTAAGATTTATTTATTTACACTATTTAAATTTTAAAAA 359
DB      546162  TTAGAT--AATTTAAGGAAAAAAGACATTTAAGTTTACCTAGCTCATATCTCATTTGA 546105

QY      360  CTAAATCACTTAATTTAATAAGACTTTGAAAATGGCCAGCGCGAGTACCTCTGCTG 419
DB      546104  AGACAGAAAAAGTGAAGTCAAAATGTTAATGAGGCCAGAGCGGCTCATGCTG 546045

QY      420  TAAATCCACACTTTGGGAGGCCCAAGTGGCGGATCACTTGAGTCAAGAGTTTAAAG 479
DB      546044  TAAATCCACAGACTTTGAGAGGCCCAAGTGGCGGATCACTTGAGTCAAGAGTTTGAAC 545985

QY      480  CAGCTGGCCCAAGATGGTGAACCCCTGCTCTACTTAAAAACGCAAAATTTAGCCAGGTG 539
DB      545984  CAGCTGGCCCAAGATGGCAAAATCCGCTCTCTACTTAAAAATCAAAATTTAGCCAGCAT 545925

QY      540  GGTGCAATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGGAAATTTGTTGAAC 599
DB      545924  GGTGGCAGGCACTTTGATCCAGCTGCTTGGGAGGCTGAGGCGAGGAAATTTGTTGAAC 545865
```



Db	546269	ATATTATTGTAATGACTTAAGTAATATTAAACTCTAGCTAAAAAGCTTCAGGGTTTC	546210
QY	240	ATTAACCTATGAAAGAGGTGTGAGATAGTGTGGTATGAATAAGATTTCAACAAGTA	299
Db	546209	TTAG-----GTGTATCTGGTTTACACACCAAGTTGAGATCACAGATTA	546163
QY	300	GTACTGCTATTGAAGATTTTAAAGATTATTATTACAATAATTATATAAAATTTTAAAA	359
Db	546162	TTAGAT--AATTTAAAGCCAAAAAAGACATTAAAGTTTACCTGACCTCAATACCTCATTTTGA	546105
QY	360	CTAATACACTTAATATTATTAAGACCTTGAAGATGGAATGAGCCAGGCGCAGTAGTCTCTGCTCG	419
Db	546104	AGACGAAAAAGTGAAGTCAAAAATGTGTTAATGAGGCGCAGAGGCGGTGCTCATGCTCG	546045
QY	420	TAAATCCACAACCTTTGGAGGCGCAAGGTGGGGATCACTGAGGTCAGAGATTAAAGC	479
Db	546044	TAAATCCAGCACTTTGAGAGGCGCAAGGTGGGGATCACTGAGGTCAGAGATTAAAGC	545985
QY	480	CAGCTGCGCAACATGGTGAACCCGTCTCTACTAAAAACCCAAAAATTAGCCAGTGT	539
Db	545984	CAGCTGCGCAACATGGCAAAATCCCGTCTCTACTAAAAATCAAAAAATTAGCCAGGAT	545925
QY	540	GGTGCATATGCACTGTAGTCCCACTACTCAGAGGTTGAGGAGGAGTAATGCTTGAAC	599
Db	545924	GGTGCAGAGCACTTGTAACTCCAGGTCTTTGGAGGCTGAGGCAAGAGTAATGCTTGAAC	545865
QY	600	CTAGAGGTGAGGTTGCAATACCCGAGAT--GTCACTGCACTCCACCT--GGCAAA	655
Db	545864	CCGTGAGGTGAGGTTGCAATACCCGAGATCAAGCACTGCACTCCACCTGGGGTAA	545805
QY	656	GAGCAAGACTCCATAAAGACAAACAAGCTTGAATTGTAAATGAGTGTAACTATC	715
Db	545804	GAGTGAAGTCCGTCTCAAAAAACAACAACCAAAACCAAAAAACATTTGAGCAATT	545745
QY	716	TTCAATTAAAGAAATTCATCTT	736
Db	545744	GGAAAGTCTGAATCTCAGCAT	545724

```

RESULT 20
US-09-949-016-14561/c
; Sequence 14561, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14561
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14561

Query Match      11.2%;   Score 223.8;   DB 3;   Length 818128;
Best Local Similarity 61.3%;   Pred. No. 4.4e-45;
Matches 454;   Conservative 0;   Mismatches 267;   Indels 20;   Gaps 5;

```

QY	1	TGATCCACGAGCCTTGCGCTCCCAAGTGTGGGATTACAGGGGTGAGCCACCACTGCTG	80
Db	546449	TGATCCACCGCGCTTGCGCTCCCAAGTGTGGGATTACAGAGTGAACCAACGCGCGC	546390
QY	61	GTGCAGTGTCTTATTAATTT-GAAGAGCAACATGGCGCTTAAATCTGTCTTATTTGA	119
Db	546389	GCGTATCATCTGTCTTCTTAAACAAGATGTAGTCTCTCGAGGGCAGGGGTTTGTCTT	546330
QY	120	CAGACTTGTATGAGTCCAAATCCCAATGCTGCCACTTACGAAACGGCGCTTAATGACTTA	179
Db	546329	GTTTGCGCTTATTAATCAAACTAAATAGCGCGCTGGCATATACAGCCCAATAG	546270
QY	180	GTCTCTCAGCGCTCTTCTTGCAATATGTAAGTGAATATATGATGTGCTTCAAGAGGA	239
Db	546269	ATATTAATGTAAATACATAAGATATTTAAACCTCCTAGTAAAAAGTCTCAGGGTTTC	546210
QY	240	ATAAACCTATGAAAAAGTTGAGATATGTGTTGATATGAATATAGATTTCAACAGTA	299
Db	546209	TTAG-----GTGTATCTGGTTTAAACACCAAGTTGAGATACAGAAATTA	546163
QY	300	GTAGCTGTATGTAAGATTTTAAAGTTATTTATTAACAATTAATTAATTTTAAAAA	359
Db	546162	TTAGAT--AATTTAAAGCGAAAAAGACATTTAAGTTTACTTGCTCATATCCTATTTGA	546105
QY	360	CTAATACACTTAATTTATTTAAAGCTTTGAAATGGGCGAGGCGAGTAGCTCTGCTG	419
Db	546104	AGACGAAAAAGTGAATCTCAAAATGTTAATAGGCGCAGAGCGGTGCTCAGCTCG	546045
QY	420	TAATCCAAACTTTGGAGGCGCAAGTGGCGGATACCTGAGTCCAGAGTTTAAAGC	479
Db	546044	TAATCCCAAGCTTTAGAGAGGCGAAGTGGCGGATCACTGAGGTCCAGAACTTGAAGC	545985
QY	480	CAGCTTGCCAAACAGTGAACCTGTCTCTACTTAAAAACGCAAAATTTAGCCAGTGT	539
Db	545984	CAGCTTGACCAACATGGCAAAATCCCGTCTCTACTTAAAAATACAAAAATTTAGCCAGCAT	545925
QY	540	GGTGGCATACGACTGTAGTCCCACTACTCTCAGAGGTTGAGGAGAGAAATGCTTTGAAC	599
Db	545924	GGTGGCAGGCACTTGTAACTCCAGTGTGTTGAGAGGCTGAGCGAGAAATGCTTTGAAC	545865
QY	600	CTAGAGGTGAGGTTTGCAATCCCGAGAT--GTCACTGCACTTCAGCCT-GGCAACA	655
Db	545864	CCGTAGGGTGGAGTTTGCAGTAGCCCAAGATACGCCCACTGCACTTCACGCGGGGGA	545805
QY	656	GAGCAAGATTCATTAAGCAACAAGCTTTGAATGTGTAAATGAGTGTACCTATC	715
Db	545804	GAGTGACTCCGTCTCAAAAAACAACAACAACAACAACAATTTGAGACATTT	545745
QY	716	TTCAATTAGAATTCATCTT736	
Db	545744	GGAGATCTGAATCTCAGCAT545724	

```

RESULT 21
US-09-949-016-14562/c
; Sequence 14562, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14562

```

```
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14562

Query Match      11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4,4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY      1  TGAATCACCAGCGCTGGCCCTCCCAAGTGTGGAGTTATGAGGGGTGAGCCACCAAGCCTG 60
DB      546449  TGAATCACCAGCGCTGGCCCTCCCAAGTGTGGAGTTATGAGGGGTGAGCCACCAAGCCTG 546390
QY      61  GTGCAATGCTTATTATT- GAAGAGACAATGGGCTTAATCTGCTCTATTGGA 119
DB      546389  GCCTATCATCTGCTCTTCTTACAGAATGTAAGCTCTCCGAGGGGAGGGGCTTTGCTT 546330
QY      120  CAGACTTTGATGAGTCAAAATCCCAATGCTGCCACTTACTGAAAGGCTTAAATGACTTA 179
DB      546329  GTTGTGCTGCTTTATTATCAATACCTAAATGGCGCTGGCATATTAACAGACCAATATG 546270
QY      180  GTCTCTCAGCTGCTCTTCTGCAATGTAAGTGAATATGATGCTTTCAAGAGGA 239
DB      546269  ATATTATATGAATGACTAAGTGAATATTTAACTCCTAGCTAAAGTCTTCAGGGCTTC 546210
QY      240  ATAACTATGAAAGTGTGAGATAGTGTGATGATGTAATTAAGATTCAACAAGTA 299
DB      546209  TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAGAAATCAAGATTA 546163
QY      300  GTAGCTGCTATTGAAGATTTAAGATTATTATTATTAACAATTAAATTTTAAAA 359
DB      546162  TTAGAT--AATTAAAGCAAAAAAGACATTAACTTAAGTTAACTCAATACCTCATTTTGA 546105
QY      360  CTAAATACCTTAATTTATTAAGCTTTGAATGGGCGCGGCACTAGTCTCTGCTG 419
DB      546104  AGCAGAAAAAGTGAAGTCAAAATGTTAAATGAGCGCGAGCGGTGCTCATGCTG 546045
QY      420  TAATCCCAACACTTTGGAGGCGCAAGTGGGCGGATCACTGAGTCAAGAGTTTAAG 479
DB      546044  TAATCCCAACACTTTGGAGGCGCAAGTGGGCGGATCACTGAGTCAAGAGTTTAAG 545985
QY      480  CAGCTGGCCCAATGCTGTAACCTGCTCTTAATAAAACGCAAAATTTAGCCAGGT 539
DB      545984  CAGCTGACCAACATGCAAAATCCGCTCTTAATAAAATACAAAAATTTAGCCAGGT 545925
QY      540  GGTGGCATGCACTGTATGTCCTCACTACAGAGGTTGAGGAGAGAAATGCTTGAAC 599
DB      545924  GGTGGCAGGCACTGTATGTCCTCACTACAGAGGTTGAGGAGAGAAATGCTTGAAC 545865
QY      600  CTAGAGGTGGAGGTTGAGTAACCGAAGT--GTCACTGCACTCAGGCT--GGCAAC 655
DB      545864  CCGTGGGTGGAGGTTGAGTGAAGTGAAGTCAAGCACTGCACTCAGGCTGGGTGAAC 545805
QY      656  GAGCAAGACTCCATAAAGCAAAAGCTTTGAAATGTTGTAATGAGTTGTAATCTAT 715
DB      545804  GAGTGAAGCTCCGTCTCAAAAAACAAAAACAAAAACAAAAACATTTGAGACATT 545745
QY      716  TTCATTTAAGAAATTCATCTT 736
DB      545744  GGAAGATCTGAATCTCAGCAT 545724

RESULT 22
US-09-949-016-14564/C
; Sequence 14564, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14564
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14564

Query Match      11.2%; Score 223.8; DB 3; Length 818128;
Best Local Similarity 61.3%; Pred. No. 4,4e-45;
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY      1  TGAATCACCAGCGCTGGCCCTCCCAAGTGTGGAGTTATGAGGGGTGAGCCACCAAGCCTG 60
DB      546449  TGAATCACCAGCGCTGGCCCTCCCAAGTGTGGAGTTATGAGGGGTGAGCCACCAAGCCTG 546390
QY      61  GTGCAATGCTTATTATT- GAAGAGACAATGGGCTTAATCTGCTCTATTGGA 119
DB      546389  GCCTATCATCTGCTCTTCTTACAGAATGTAAGCTCTCCGAGGGGAGGGGCTTTGCTT 546330
QY      120  CAGACTTTGATGAGTCAAAATCCCAATGCTGCCACTTACTGAAAGGCTTAAATGACTTA 179
DB      546329  GTTGTGCTGCTTTATTATCAATACCTAAATGGCGCTGGCATATTAACAGACCAATATG 546270
QY      180  GTCTCTCAGCTGCTCTTCTGCAATGTAAGTGAATATGATGCTTTCAAGAGGA 239
DB      546269  ATATTATATGAATGACTAAGTGAATATTTAACTCCTAGCTAAAGTCTTCAGGGCTTC 546210
QY      240  ATAACTATGAAAGTGTGAGATAGTGTGATGATGTAATTAAGATTCAACAAGTA 299
DB      546209  TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAGAAATCAACAAGATTA 546163
QY      300  GTAGCTGCTATTGAAGATTTAAGATTATTATTATTAACAATTAAATTTTAAAA 359
DB      546162  TTAGAT--AATTAAAGCAAAAAAGACATTAACTTAAGTTAACTCAATACCTCATTTTGA 546105
QY      360  CTAAATACCTTAATTTATTAAGCTTTGAATGGGCGCGCAAGGCACTAGTCTCTGCTG 419
DB      546104  AGCAGAAAAAGTGAAGTCAAAATGTTAAATGAGCGCGAGCGGTGCTCATGCTG 546045
QY      420  TAATCCCAACACTTTGGAGGCGCAAGTGGGCGGATCACTGAGTCAAGAGTTTAAG 479
DB      546044  TAATCCCAACACTTTGGAGGCGCAAGTGGGCGGATCACTGAGTCAAGAGTTTAAG 545985
QY      480  CAGCTGGCCCAATGCTGTAACCTGCTCTTAATAAAACGCAAAATTTAGCCAGGT 539
DB      545984  CAGCTGACCAACATGCAAAATCCGCTCTTAATAAAATACAAAAATTTAGCCAGGT 545925
QY      540  GGTGGCATGCACTGTATGTCCTCACTACAGAGGTTGAGGAGAGAAATGCTTGAAC 599
DB      545924  GGTGGCAGGCACTGTATGTCCTCACTACAGAGGTTGAGGAGAGAAATGCTTGAAC 545865
QY      600  CTAGAGGTGGAGGTTGAGTAACCGAAGT--GTCACTGCACTCAGGCT--GGCAAC 655
DB      545864  CCGTGGGTGGAGGTTGAGTGAAGTGAAGTCAAGCACTGCACTCAGGCTGGGTGAAC 545805
QY      656  GAGCAAGACTCCATAAAGCAAAAGCTTTGAAATGTTGTAATGAGTTGTAATCTAT 715
DB      545804  GAGTGAAGCTCCGTCTCAAAAAACAAAAACAAAAACAAAAACATTTGAGACATT 545745
```

QY 716 TTCATTAGAAATTCATCTT 736  
Db 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 23  
US-09-949-016-14565/C  
; Sequence 14565, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14565  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14565

Query Match 11.2%; Score 223.8; DB 3; Length 818128;  
Best Local Similarity 61.3%; Pred. No. 4.4e-45;  
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY 1 TGATCCACGACCTTGCGCTCCCAAGTGTGGATTACAGGCGTGACCAACGCGCTG 60  
Db 546449 TGATCCACCGCTTGCGCTCCCAAGTGTGGATTACAGGCGTGACCAACGCGCTG 546390

QY 61 GTCGAATGCTTTATTTT-GAAGACAAACATGGCCCTTAATCTGCTTATTTGA 119  
Db 546389 GCTTATATCTGCTTTTACAAAGATGTAAGTCTCTCGAAGGCGAGGGTTTGTCTT 546330

QY 120 CAGACTTGAATGAGTCAATCCCAATGCTGCACTTACTGAAACGCGCTTAATGACTTA 179  
Db 546329 GTTTGCTGCTTATTTATCAATACCTAAATGCGCGCTGCATATACAGACCCACATAG 546270

QY 180 GTCTCTCAGCTGCTTTTGTGCAATGTAAAGTGAATATGATGCTTTCAAGAGGA 239  
Db 546269 ATATTAATTAAGTACTTAAGTAATTTAAATCTAGCTAAAGCTTCAGGGTTTC 546210

QY 240 ATAAACCTATGAAGAAGTGAAGATGTTGATTAAGAAATAGAGATTCAACAAGTA 299  
Db 546209 TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAAGAAATACGAATTA 546163

QY 300 GTAAGCTATTAAGATTAAAGATTATTTATTAACAACCTAATTAATTAATTTAAAAA 359  
Db 546162 TTAAAT--AATTAAGGAAAAAAGACATTAAGTTTACCTAGCTCAATACCTCAATTTGA 546105

QY 360 CTAATACACTTAATTTATTAAGACTTTGAAAAGGCGGACGCAAGTACTCTGCTG 419  
Db 546104 AGACAGAAAAAGTGAAGTCAAAATGTTAAATAGAGGCGCAGAGCGGTCAATGCTG 546045

QY 420 TAAATCCAAACCTTTGGAGGCGCAAGTGGGCGGATACCTGAGGTGAGAGTTAAAGAC 479  
Db 546044 TAAATCCAAACCTTTAGAGGCGCAAGTGGGCGGATACCTGAGGTGAGAGTTTGAAC 545985

QY 480 CAGCTGCGCAACATGTGAACCTGTCTCTACTAAAAACGCAAAATTAATGCAAGGTGT 539  
Db 545985 CAGCTGCGCAACATGTGAACCTGTCTCTACTAAAAACGCAAAATTAATGCAAGGTGT 539

Db 545984 CAGCTGACCAACATGCGAAATCCCGTCTCTACTAAAAATACAAAAATAGCAGGCAT 545925  
QY 540 GTGGCATGACCTGTAGTCCCACTACAGAGGTTGAGGAGAGATTGTTGAAC 599  
Db 545924 GTGGCATGACCTGTAGTCCCACTACAGAGGTTGAGGAGAGATTGTTGAAC 545865

QY 600 CTAGAGGTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTT 655  
Db 545864 CCGTGAAGTGGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTT 545805

QY 656 GAGCAAGACTCCATTAAGACAAAGCTTTGAATTTGTAATGATGTTGATGTTATC 715  
Db 545804 GAGTGAAGTCCGTCTCAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 545745

QY 716 TTCATTAGAAATTCATCTT 736  
Db 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 24  
US-09-949-016-14566/C  
; Sequence 14566, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14566  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14566

Query Match 11.2%; Score 223.8; DB 3; Length 818128;  
Best Local Similarity 61.3%; Pred. No. 4.4e-45;  
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;

QY 1 TGATCCACGACCTTGCGCTCCCAAGTGTGGATTACAGGCGTGACCAACGCGCTG 60  
Db 546449 TGATCCACCGCTTGCGCTCCCAAGTGTGGATTACAGGCGTGACCAACGCGCTG 546390

QY 61 GTCGAATGCTTTATTTT-GAAGACAAACATGGCCCTTAATCTGCTTATTTGA 119  
Db 546389 GCTTATATCTGCTTTTACAAAGATGTAAGTCTCTCGAAGGCGAGGGTTTGTCTT 546330

QY 120 CAGACTTGAATGAGTCAATCCCAATGCTGCACTTACTGAAACGCGCTTAATGACTTA 179  
Db 546329 GTTTGCTGCTTATTTATCAATACCTAAATGCGCGCTGCATATACAGACCCACATAG 546270

QY 180 GTCTCTCAGCTGCTTTTGTGCAATGTAAAGTGAATATGATGCTTTCAAGAGGA 239  
Db 546269 ATATTAATTAAGTACTTAAGTAATTTAAATCTAGCTAAAGCTTCAGGGTTTC 546210

QY 240 ATAAACCTATGAAGAAGTGAAGATGTTGATTAAGAAATAGAGATTCAACAAGTA 299  
Db 546209 TTAG-----GTGTAATCTGTTTAAACACCAAGTTGAAGAAATACGAATTA 546163

QY 300 GTAAGCTATTAAGATTAAAGATTATTTATTAACAACCTAATTAATTAATTTAAAAA 359  
Db 545985 CAGCTGCGCAACATGTGAACCTGTCTCTACTAAAAACGCAAAATTAATGCAAGGTGT 539

Db 546162 TTAGAT--AATTAAAGCAAAAAGACATTAAGTTTAACTAGCTCATATACCTTATTTGA 546105  
Qy 360 CTATACACTTAATTAATTTAAAGACTTTGAATGGCCGCGCAGTACTCTGCTG 419  
Db 546104 AGCAGAAAAAGTAAAGTCAAAATGTTAAATGAGGCCGAGCGGTGCTCATGCTG 546045  
Qy 420 TAATCCCAACACTTTGGGAGGCCCAAGTGGCGGATCACCTGAGTCAAGAGTTTAAAG 479  
Db 546044 TAATCCCAACACTTTGGGAGGCCCAAGTGGCGGATCACCTGAGTCAAGAGTTTAAAG 545985  
Qy 480 CAGCTGCGCAACATGCTGTAAGACCTGCTCTTAATAAAGCAAAATTTAGCCAGGTG 539  
Db 545984 CAGCTGCGCAACATGCTGTAAGACCTGCTCTTAATAAAGCAAAATTTAGCCAGGTG 545925  
Qy 540 GGTGGCATGCACTGTATGTCCTCACTACTCAGAGAGTTGAGGAGAAATTCCTTGAC 599  
Db 545924 GGTGGCATGCACTGTATGTCCTCACTACTCAGAGAGTTGAGGAGAAATTCCTTGAC 545865  
Qy 600 CTAGAGAGTGGAGGTTGACGTAAACCCGAGAT--GTCACTGCACCTCAGCT--GGCAACA 655  
Db 545864 CCGTAGAGTGGAGGTTGACGTAAACCCGAGAT--GTCACTGCACCTCAGCT--GGCAACA 545805  
Qy 656 GAGCAAGACTCCATTAAGAACAACAAGCTTTGAATTTGTAAATGAGTTTACTTATC 715  
Db 545804 GAGTGAAGTCTCCGTCTCAAAAACAAACAAACAAACAAACAAACAAACAAACAAAC 545745  
Qy 716 TTCATTAAAGAAATTCATCTT 736  
Db 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 25  
US-09-949-016-14567/c  
; Sequence 14567, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 14567  
; LENGTH: 818128  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(818128)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14567

Query Match 11.2%; Score 223.8; DB 3; Length 818128;  
Best Local Similarity 61.3%; Pred. No. 4,4e-45;  
Matches 454; Conservative 0; Mismatches 267; Indels 20; Gaps 5;  
Qy 1 TGATCCACGAGCTTGGCTCCCAAGTCTGGAGTTAAGAGCGGTAGCCACGCGCTG 60  
Db 546449 TGATCCACGAGCTTGGCTCCCAAGTCTGGAGTTAAGAGCGGTAGCCACGCGCG 546390  
Qy 61 GTCGAATGCTTTAATTAATTT-GAAGAGACAACATGGGCTTAAATCTGCTCTTATTTGA 119  
Db 546389 GCTATCATCTGCTCTTATTAACAAGATGTAAGCTCTCCTGAGGGGAGGTTTGTCTT 546330

Qy 120 CAGACTTGTATGAGCAATATCCCAATGCTGCACCTTACTGAACGCGCTTAATAGACTTA 179  
Db 546329 GTTGTGCTGCTTATTTCAATATCTAAATATGGCGCTTGCGCATTAACAGCCACATAG 546270  
Qy 180 GTCTCTTCAGCTGTCTTCTGCAATATGTAAGTGAATATATGATGCTTTCAAGAGGA 239  
Db 546269 ATATTAATGAATGACTAAGTGAATATTAATTAATCTAGCTAAAGTCTTCAGGGTTTC 546210  
Qy 240 ATTAACCTATGAAGAGTGTGAGATATGTTGATATGTAATATAGATTTCAACAAGTA 239  
Db 546209 TTAG-----GTATATCTGCTTTAACAACAAGTGAAGATTCACAGATTA 546163  
Qy 300 GTAGCTGTATTAAGATTTAAGATTATTTAATTAACATTAATTAATTAATTAATTA 359  
Db 546162 TTAGAT--AATTAAAGCAAAAAGCAATTAAGTTTAACTAGCTCAATATCTATTTGA 546105  
Qy 360 CTATACACTTAATTAATTTAAAGAGCTTTGAATGGGCCAGGCGAGTACTCTGCTG 419  
Db 546104 AGCAGAAAAAGTAAAGTCAAAATGTTAAATGAGGCCAGGCGGTGCTCATGCTG 546045  
Qy 420 TAATCCCAACACTTTGGGAGGCCCAAGTGGCGGATCACCTGAGTCAAGAGTTTAAAG 479  
Db 546044 TAATCCCAACACTTTGGGAGGCCCAAGTGGCGGATCACCTGAGTCAAGAGTTTAAAG 545985  
Qy 480 CAGCTGCGCAACATGCTGTAAGACCTGCTCTTAATAAAGCAAAATTTAGCCAGGTG 539  
Db 545984 CAGCTGCGCAACATGCTGTAAGACCTGCTCTTAATAAAGCAAAATTTAGCCAGGTG 545925  
Qy 540 GGTGGCATGCACTGTATGTCCTCACTACTCAGAGGTTGAGGAGAGAAATTCCTTGAC 599  
Db 545924 GGTGGCATGCACTGTATGTCCTCACTACTCAGAGGTTGAGGAGAGAAATTCCTTGAC 545865  
Qy 600 CTAGAGAGTGGAGGTTGACGTAAACCCGAGAT--GTCACTGCACCTCAGCT--GGCAACA 655  
Db 545864 CCGTAGAGTGGAGGTTGACGTAAACCCGAGAT--GTCACTGCACCTCAGCT--GGCAACA 545805  
Qy 656 GAGCAAGACTCCATTAAGAACAACAAGCTTTGAATTTGTAAATGAGTTTACTTATC 715  
Db 545804 GAGTGAAGTCTCCGTCTCAAAAACAAACAAACAAACAAACAAACAAACAAACAAAC 545745  
Qy 716 TTCATTAAAGAAATTCATCTT 736  
Db 545744 GGAAGATCTGAATCTCAGCAT 545724

RESULT 26  
US-09-949-016-11957/c  
; Sequence 11957, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 11957  
; LENGTH: 64813  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11957

Query Match 10.9%; Score 217.6; DB 3; Length 64813;  
Best Local Similarity 59.7%; Pred. No. 6,4e-44;  
Matches 426; Conservative 0; Mismatches 274; Indels 14; Gaps 3;



US-09-949-016-16064/c  
; Sequence 16064, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16064  
; LENGTH: 70131  
; TYPE: DNA  
; ORGANISM: Human

1 TGATCCACAGCCTTGGCCTCCCAAGTCTGGGATTAACAGCGCTGAGCCACAGCCCTG 60  
57787 TGATCCACAGCCTTGGCCTCCCAAGTCTGGGATTAACAGCGCTGAGCCACAGCCCA 57728  
61 GTCCGATGCTTATTTATTTGAAGACAAATGGGCTTAAATCTGCTCTTATTTGAC 120  
57727 GCTCATTTCTACATTTCTGATTTAGCTTACAAATGTCATGCTCTTGAACCCAGAAA 57668  
121 AGACTTGAAGAGTCAATCCCAATGCTGCACTTAAGCGGCTTAAATGAAGTTAG 180  
57667 TTCTACTTATTAATAATACAGGTCAAGAAAATACCAAGATGTAATAAAACATTTATGTAC 57608  
181 TCTCTGAGCTGTCTTCTGCAATATGTAAGTGAATTAATAGCTTCTCAAGAGGAA 240  
57607 TATTTACCATATTAATTTAAACCAATGAAATGGAAATGGCAATTAAGCTAACATGCT 57548  
241 TAAACCTATGAAGAGTTGAGGATAGTGTGATTAATGAATTAAGATTTCAACAAGTAG 300  
57547 TCCCTGAATTAATGGGTAATTAATGCTTTGTCTGTAGTAAATTAATCTCATCGGCCG 57488  
301 TAGCTGCTATTAAGATTTAAGATTAATTAATTAACAATTAATTAATTAATTAATTA 360  
57487 TAAATATCATATTAATTAAGATTAATCAATTAAGAAATGTTCTCGGT----- 57436  
361 TAAATCATTAATTAATTAAGATTTGAAATGGGCCAGGCGCAATAGCTCTGCTGT 420  
57435 --ATATGTTAGGTAGAAAGAAATCAAGTTACGGCCAGGCAAGTGTCTACACCTAT 57378  
421 AATCCCAACACTTTGGGAGGCGCAAGGTGGCGGATCACTGAGGTGAGGATTTAAGACC 480  
57377 AATCCCAACACTTTGGGAGGCGTGAAGGCGGATCACTGAGGTGAGGATTTGAGACC 57318  
481 AGCTGCGCAACATGTGTGAACCCCTGTCTCTACTAAACGCAAAATTAAGCAAGGTGTG 540  
57317 AGCTGCGCAACATGTGTGAACCCCTGTCTCTACTAAACGCAAAATTAAGCAAGGACG 57258  
541 GTGCGATGCACTGTGATGCCAATCTACTGAGAGTGTGAGGAGGAGAAATGCTTTAACC 600  
57257 GTGCGATGCGCTGTGTGCTCAGCTACTGAGAGGCTGAGGAGGAGAAATGCTTTAACC 57198  
601 TAGGAGGTGAGGTTGAGTAACCCGAG--ATGTCACCTGACTCCAGCCTGG-CAACAG 656  
57197 TGGAGGTGAGGTTTACTGAGCGGAGATCATGCCCTGCACTCCAGCCTGAGTAGCAG 57138  
657 AGCAAGACTCCATTAAGACAAACAAAGCTTTGAATTTGTGTAATGAAGTTGTAC 710  
57137 AGTGAAGCTCATTAATAAAAAAAAAAAGTTTCAAAATTTATGCAAAATTTATGAC 57084

US-09-949-016-16064  
Query Match 10.9%; Score 217.6; DB 3; Length 70131;  
Best Local Similarity 59.7%; Pred. No. 6,5e-44;  
Matches 426; Conservative 0; Mismatches 274; Indels 14; Gaps 3;

1 TGATCCACAGCCTTGGCCTCCCAAGTCTGGGATTAACAGCGCTGAGCCACAGCCCTG 60  
63441 TGATCCACAGCCTTGGCCTCCCAAGTCTGGGATTAACAGCGCTGAGCCACAGCCCA 63382  
61 GTCCGATGCTTATTTATTTGAAGACAAATGGGCTTAAATCTGCTCTTATTTGAC 120  
63381 GCTCATTTCTACATTTCTGATTTAGCTTACAAATGTCATGCTCTTGAACCCAGAAA 63322  
121 AGACTTGAAGAGTCAATCCCAATGCTGCACTTAAGCGGCTTAAATGAAGTTAG 180  
63321 TTCTACTTATTAATAATACAGGTCAAGAAAATACCAAGATGTAATAAAACATTTATGTAC 63262  
181 TCTCTGAGCTGTCTTCTGCAATATGTAAGTGAATTAATAGCTTCTCAAGAGGAA 240  
63261 TATTTACCATATTAATTTAAACCAATGAAATGGAAATGGCAATTAAGCTAACATGCT 63202  
241 TAAACCTATGAAGAGTTGAGGATAGTGTGATTAATGAATTAAGATTTCAACAAGTAG 300  
63201 TCCCTGAATTAATGGGTAATTAATGCTTTGTCTGTAGTAAATTAATCTCATCGGCCG 63142  
301 TAGCTGCTATTAAGATTTAAGATTAATTAATTAACAATTAATTAATTAATTAATTA 360  
63141 TAAATATCATATTAATTAAGATTAATCAATTAAGAAATGTTCTCGGT----- 63090  
361 TAAATCATTAATTAATTAAGATTTGAAATGGGCCAGGCGCAATAGCTCTGCTGT 420  
63089 --ATATGTTAGGTAGAAAGAAATCAAGTTACGGCCAGGCAAGTGTCTACACCTAT 63032  
421 AATCCCAACACTTTGGGAGGCGCAAGGTGGCGGATCACTGAGGTGAGGATTTAAGACC 480  
63031 AATCCCAACACTTTGGGAGGCGTGAAGGCGGATCACTGAGGTGAGGATTTGAGACC 62972  
481 AGCTGCGCAACATGTGTGAACCCCTGTCTCTACTAAACGCAAAATTAAGCAAGGTGTG 540  
62971 AGCTGCGCAACATGTGTGAACCCCTGTCTCTACTAAACGCAAAATTAAGCAAGGACG 62912  
541 GTGCGATGCACTGTGATGCCAATCTACTGAGAGTGTGAGGAGGAGAAATGCTTTAACC 600  
62911 GTGCGATGCGCTGTGTGCTCAGCTACTGAGAGGCTGAGGAGGAGAAATGCTTTAACC 62852  
601 TAGGAGGTGAGGTTGAGTAACCCGAG--ATGTCACCTGACTCCAGCCTGG-CAACAG 656  
62851 TGGAGGTGAGGTTTACTGAGCGGAGATCATGCCCTGCACTCCAGCCTGAGTAGCAG 62792  
657 AGCAAGACTCCATTAAGACAAACAAAGCTTTGAATTTGTGTAATGAAGTTGTAC 710  
62791 AGTGAAGCTCATTAATAAAAAAAAAAAGTTTCAAAATTTATGCAAAATTTATGAC 62738

RESULT 28  
US-09-768-185A-1/c  
; Sequence 1, Application US/09768185A  
; Patent No. 6818758  
; GENERAL INFORMATION:  
; APPLICANT: Casseel, Michael et al  
; TITLE OF INVENTION: Estrogen receptor beta variants and  
; TITLE OF INVENTION: methods of detection thereof  
; FILE REFERENCE: C1000280  
; CURRENT APPLICATION NUMBER: US/09/768,185A  
; CURRENT FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 09768185  
; PRIOR FILING DATE: 2001-01-24  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 325791  
; TYPE: DNA

```
; ORGANISM: HUMAN
US-09-768-185A-1

Query Match      10.7%; Score 214.4; DB 3; Length 325791;
Best Local Similarity 73.0%; Pred. No. 7e-43;
Matches 305; Conservative 0; Mismatches 106; Indels 7; Gaps 2;

QY 268 TGTGTGATGATGAATAGATTAAGATTCACACAGTAGTAGTGCATTAATTAAGATTAAGATTA 327
DB 257049 TGATTGAGAGATAGATATTATTCACACAGTCTTCAACAAATTTAAATTAATTAAGATTA 256990
QY 328 TTTATTAACAATTTAATTAATTTAAATTTAAATTAATTAATTAATTAATTAATTAAGAGCTT 387
DB 256989 ATACATTTCTATGAGACATTTAATAGCAGAGGATCTTAAGAAAGTTAAGAAAGCTTT 256930
QY 388 TGAATAGGCGCAGGCGCAGTAGTCTCTGCTGTATATCCCAACACTTTGGAGGCGCAAGT 447
DB 256929 ---CCTGGCCATGCGCAGAGGCTTATGCTTGAATCCAGCACTTTGGAGGCTGAGGC 256874
QY 448 GGGGGGATGACCTGAGGTCAGAGTTTAAGACCAAGCTGGCGCAACATGTTGAACCTCTGT 507
DB 256873 GGGGTGATCACCTGAGGTCAGAGTTGAGACCAAGCTGGCGCAACATGTTGAACCTCTGT 256814
QY 508 CTCTACTAAAAAGCAAAATTAAGCAGGTGTGGTCATGCACTGTAGTCCCACTAC 567
DB 256813 CTCTACCAAAAATACAAAATTAAGCTGTGTATGTGTGCAAGCAGCTGTATCCAGCTAC 256754
QY 568 TCAGAGGTTGAGGAGAGAAATGCTTGAACCTAGAGGTGAGAGTTGAGTGAACCCGA 627
DB 256753 CTGGGAGGCTGAGGAGAGAAATGCTTGAACCTAGAGGTCGAGTTGAGTGAAGCTGA 256694
QY 628 GAT---GTACATGCTCCAGCCTGGCAAGACAGCAAGATCCATTAAGCAACAAA 682
DB 256693 GATCAGCCCTCTGACTCCAGCCTGGCGACACAGCAAGACTTATCTCAAAAAAAA 256636

RESULT 29
US-09-949-016-17578
; Sequence 17578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17578
; LENGTH: 43726
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(43726)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17578

Query Match      10.6%; Score 212.2; DB 3; Length 43726;
Best Local Similarity 81.0%; Pred. No. 1.2e-42;
Matches 260; Conservative 0; Mismatches 58; Indels 3; Gaps 1;

QY 364 TACACTTAATTAATTAAGAGCTTTGAATGGCCAGGCGCAGTAGCTCTGCTGTAAAT 423
DB 23622 TACACAAAATGTGTGAAAAATGTGAGAGCGGCGGCTGCAATGCTCAGCCTTAAT 23681
```

```
QY 424 CCCAAGCTTTGGAGGCCAAGGTGGCGGATCATCTGAGTCAAGAGTTTAAGACCAGC 483
DB 23682 CTTAGACATTTGGAGGCCAAGCGGGTGGATATCACTGAGGTCAAGAGTTTGAGCCATC 23741
QY 484 CTGGCCAAATGTTGAATACCTTGTCTCTACTAATAAAAGCAAAATTAAGCAGGTGTGTG 543
DB 23742 CTGGCCAAATGTTGAATACCTTGTCTCTAATAAAATTAAGCAGGTGTGTG 23801
QY 544 GCATGACCTGTGTCTCCAACTACTAGAGGTTGAGGAGGAATGCTTAACCTAG 603
DB 23802 GCATGACCTGTGTCTCCAACTACTAGAGGTTGAGGAGGAATGCTTAACCTAG 23861
QY 604 GAGTGAAGTTGTCAGTAAACCCGAGAT---GTCACTGACTCTCAGCCTGGCAACAGCA 660
DB 23862 GAGTGAAGTTGTCAGTAAACCCGAGATCGGCGCATTTGATTCAGCCTGGCAACAGCA 23921
QY 661 AGACTCCATTAAGACAAACA 681
DB 23922 AGACTCATCTCAAAAAATTA 23942

RESULT 30
US-09-949-016-17482
; Sequence 17482, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17482
; LENGTH: 76962
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(76962)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17482

Query Match      10.6%; Score 212.2; DB 3; Length 76962;
Best Local Similarity 78.9%; Pred. No. 1.5e-42;
Matches 266; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

QY 378 TAAAGCTTTGAATAGGCGCAGGCGCAGTAGCTCTGCTGTAAATCCCAACCTTTGG 437
DB 45917 TTAAGAGCAGACCTGTGTGCGCGGCGCAGTGTCTACGCTGTAAATCCCAACCTTTGG 45976
QY 438 AGGCCAAGGTGGCGGATCATCTGAGTCAAGAGTTTAAGACAGCCTGGCCAACTGTG 497
DB 45977 AGGCCAAGGTGGCGGATCATCTGAGTCAAGAGTTTACAGCAGCCTGGCCAACTGTG 46036
QY 498 GAAACCTGTCTCTACTAATAAAAGCAAAATTAAGCAGGTGTGTGTCATGCACTGTAG 557
DB 46037 GAAACCTGTCTCTACTAATAAAAGCAAAATTAAGCAGGTGTGTGTCATGCACTGTAG 46096
QY 558 TCCCACTACTCAGAGGTTGAGGAGGAATGCTTAACCTAGAGTGTGAGGTTGC 617
DB 46097 TCCCACTACTCAGAGGTTGAGGAGGAATGCTTAACCTAGAGGTTGTG 46156
QY 618 AGTAACCCGAG---ATGTCACTGCACTCCAGCCTGGCAACAGACAAAGACTCCATTAAG 674
DB 46157 AGTAACCCGAG---ATGTCACTGCACTCCAGCCTGGGTAACAGAGAAAGACTCCATTAAG 46216
```

OY 675 CAACAAAGCTTGAATGTGTAAATGAGTTGACC 711  
Db 46217 AAAAAAAAAAAAAAAAAAAAAAAAAAGGAGAGCC 46253

RESULT 31  
US-09-949-016-12550  
; Sequence 12550, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12550  
; LENGTH: 45842  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12550

Query Match 10.6%; Score 211.8; DB 3; Length 45842;  
Best Local Similarity 65.0%; Pred. No. 1.6e-42;

Matches 379; Conservative 0; Mismatches 192; Indels 12; Gaps 4;

OY 259 TGAGATAGTGTGATGATGAATTAAGATTTCACAAAGTAGTAGCTGTATGGAAGATT 318  
Db 23739 TAAACAAATATGTTCTAATATATTAAGTAATTAATATATGAATGAATTAATTCACAA 23798  
OY 319 TAAAGTATTTATTAACAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 375  
Db 23799 ATCAAAAAATATATATCAAGTTGTATATATGATGCTTACAAAGAAACCAACTTAATATGAT 23858  
OY 376 --ATTAAAGAGCTTTGAAATGGGCGAGGCGAGTAGCTCTGCTGTATATCCCAACTT 433  
Db 23859 ACATTAATAAAATGAAAAAAGGCGAGTCCAGTGGCTCACACTTGTATATCCAGCACTT 23918  
OY 434 TGGAGGCGCAAGGTGGGCGGATCACTGAGGTCAAGAGTTAAGACCAAGCTGGCCACA 493  
Db 23919 TGGAAAGCCAAAGGTGGGCGGATCATCTGAGGTCAAGAGTTGAAACCAAGCTGGCCAAATG 23978  
OY 494 TGGTGAACCCCTGTCTCTACTTAATAAAGCAAAATTAAGCCAGGTGTGGCATGACCT 553  
Db 23979 TGGTGAACCTGTCTCTCTACTTAATAAATTAAGCCAGGTGTGGTGGGCGCT 24038  
OY 554 GTAGTCCCACTACTGAGAGGTTGAGGAGAGAAATTTGTAACCTAGGAGGTGGAGG 613  
Db 24039 ATATATCCCGCTACTCTGAGAGGCTGAGGCGAGAAATTTGTAACCTGGGAGCAAGG 24098  
OY 614 TTGCAGTAACCCGAG--ATGTCACTGACTCCAGCTT-GGCAACAGAGCAAGACTCCAT 669  
Db 24099 TTGCAGTAGGCGGAGATCATGCGACTGCACTCCAGCTGGGCGGAGCAAGATGCGCT 24158  
OY 670 ---AAAGACAACAAAGCTTTGAAATTTGTAATTAAGAGTTTACTTCACTTAAAGA 726  
Db 24159 CTCAAAAAGAGAACCAACAAACCAAAACCTGAAAAAGCAATTCGTTTTTATTTCTGGGAA 24218  
OY 727 AATTCACTTTGTTCATTATTTTACTTGACATGAGAGCTTCCAGCAATTTTAAATTA 786  
Db 24219 AGCTGCTGGCTCTTTTCCCAAGATCAAGAGTGAACAATATAGAAATTAATTAATTA 24278  
OY 787 GCCCTCAGATTTTATGTCTACCTGGCTATGTGATTAACAAAT 829

Db 24279 ACCCTTATGARTTGTGCTGGGAGGAGATATTTGAACCTGTT 24321

RESULT 32  
US-09-949-016-17327  
; Sequence 17327, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17327  
; LENGTH: 45842  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17327

Query Match 10.6%; Score 211.8; DB 3; Length 45842;  
Best Local Similarity 65.0%; Pred. No. 1.6e-42;

Matches 379; Conservative 0; Mismatches 192; Indels 12; Gaps 4;

OY 259 TGAGATAGTGTGATGATGAATTAAGATTTCACAAAGTAGTAGCTGTATGGAAGATT 318  
Db 23739 TAAACAAATATGTTCTAATATATTAAGTAATTAATATATGAATGAATTAATTCACAA 23798  
OY 319 TAAAGTATTTATTAACAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 375  
Db 23799 ATCAAAAAATATATATCAAGTTGTATATATGATGCTTACAAAGAAACCAACTTAATATGAT 23858  
OY 376 --ATTAAAGAGCTTTGAAATGGGCGAGGCGAGTAGCTCTGCTGTATATCCCAACTT 433  
Db 23859 ACATTAATAAAATGAAAAAAGGCGAGTCCAGTGGCTCACACTTGTATATCCAGCACTT 23918  
OY 434 TGGAGGCGCAAGGTGGGCGGATCACTGAGGTCAAGAGTTTAAAGCAAGCTGGCCACA 493  
Db 23919 TGGAAAGCCAAAGGTGGGCGGATCATCTGAGGTCAAGAGTTGAAACCAAGCTGGCCAAATG 23978  
OY 494 TGGTGAACCCCTGTCTCTACTTAATAAAGCAAAATTAAGCCAGGTGTGGCATGACCT 553  
Db 23979 TGGTGAACCTGTCTCTCTACTTAATAAATTAAGCCAGGTGTGGTGGGCGCT 24038  
OY 554 GTAGTCCCACTACTGAGAGGTTGAGGAGAGAAATTTGTAACCTAGGAGGTGGAGG 613  
Db 24039 ATATATCCCGCTACTCTGAGAGGCTGAGGCGAGAAATTTGTAACCTGGGAGCAAGG 24098  
OY 614 TTGCAGTAACCCGAG--ATGTCACTGACTCCAGCTT-GGCAACAGAGCAAGACTCCAT 669  
Db 24099 TTGCAGTAGGCGGAGATCATGCGACTGCACTCCAGCTGGGCGGAGCAAGATGCGCT 24158  
OY 670 ---AAAGACAACAAAGCTTTGAAATTTGTAATTAAGAGTTTACTTCACTTAAAGA 726  
Db 24159 CTCAAAAAGAGAACCAACAAACCAAAACCTGAAAAAGCAATTCGTTTTTATTTCTGGGAA 24218  
OY 727 AATTCACTTTGTTCATTATTTTACTTGACATGAGAGCTTCCAGCAATTTTAAATTA 786  
Db 24219 AGCTGCTGGCTCTTTTCCCAAGATCAAGAGTGAACAATATAGAAATTAATTAATTA 24278  
OY 787 GCCCTCAGATTTTATGTCTACCTGGCTATGTGATTAACAAAT 829  
Db 24279 ACCCTTATGARTTGTGCTGGGAGGAGATATTTGAACCTGTT 24321

```
RESULT 33
US-09-949-016-12731/c
; Sequence 12731, Application US/09949016
; Patent No. 6812339
; ORGANISM: Human
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12731
; LENGTH: 264206
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12731

Query Match
Best Local Similarity 10.6%; Score 211.6; DB 3; Length 264206;
Matches 274; Conservative 0; Mismatches 64; Indels 4; Gaps 2;

QY 372 AATTATTAAGAGCTTTGAATGGCCAGGCGAGTACTCTGCTGTAATCCCAAC 431
DB 60222 AATCTTAAGTCACTTAATAGCCGGGACAGTGGCTCAACCTGTATCCGAC 60163
QY 432 TTGGGAGGCCAAGTGGGGGATCACTGAGTCAAGATTAAAGCCAGCTGGCCAA 491
DB 60162 TTGGGAGGCCAAGGCGGGGATCACTGAGATCAAGATTGAGACCGCTGGCCAA 60103
QY 492 CATGTGAACCTGTCTCTACTAATAAAGCAAAATTTGCGGAGTGGTGGCATGCAC 551
DB 60102 CATGTGAACCTGTCTCTAATAAATAAATAAATTAAGCCAGATGTGTGGCAGGTGC 60043
QY 552 CTGTAGTCCCAACTACTCTAGAGGTTGAGGAGAGAAATGCTTGAACCTAGAGGTGA 611
DB 60042 CTGTATCCCAAGTACTCGGAGGCTGAGGTTGAGGAGATGCTTGAAGCTGGGGGCGCA 59983
QY 612 GGTTCAGTACCCTGAG--ATGTCACTGCACTCCAGCTT-GGCAACAGAGCAATCC 667
DB 59982 GGTTCAGTACCCTGAGCAGATGCACTGCACTCCAGCTGGGCAACAGAGCAATCC 59923
QY 668 ATTAAGACAACAAGCTTTGAATTTGTGAATGAGTTGA 709
DB 59922 ATATCAAAAAAAGTCCATCTTAATAATTGA 59881

RESULT 34
US-09-949-016-13249/c
; Sequence 13249, Application US/09949016
; Patent No. 6812339
; ORGANISM: Human
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 13249
; LENGTH: 264304
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13249

Query Match
Best Local Similarity 10.6%; Score 211.6; DB 3; Length 264304;
Matches 274; Conservative 0; Mismatches 64; Indels 4; Gaps 2;

QY 372 AATTATTAAGAGCTTTGAATGGCCAGGCGAGTACTCTGCTGTAATCCCAAC 431
DB 60222 AATCTTAAGTCACTTAATAGCCGGGACAGTGGCTCAACCTGTATCCGAC 60163
QY 432 TTGGGAGGCCAAGTGGGGGATCACTGAGTCAAGATTAAAGCCAGCTGGCCAA 491
DB 60162 TTGGGAGGCCAAGGCGGGGATCACTGAGATCAAGATTGAGACCGAGCTGGCCAA 60103
QY 492 CATGTGAACCTGTCTCTACTAATAAAGCAAAATTTGCGGAGTGGTGGCATGCAC 551
DB 60102 CATGTGAACCTGTCTCTAATAAATAAATAAATTAAGCCAGATGTGTGGCAGGTGC 60043
QY 552 CTGTAGTCCCAACTACTCTAGAGGTTGAGGAGAGAAATGCTTGAACCTAGAGGTGA 611
DB 60042 CTGTATCCCAAGTACTCGGAGGCTGAGGTTGAGGAGATGCTTGAAGCTGGGGGCGCA 59983
QY 612 GGTTCAGTACCCTGAG--ATGTCACTGCACTCCAGCTT-GGCAACAGAGCAATCC 667
DB 59982 GGTTCAGTACCCTGAGCAGATGCACTGCACTCCAGCTGGGCAACAGAGCAATCC 59923
QY 668 ATTAAGACAACAAGCTTTGAATTTGTGAATGAGTTGA 709
DB 59922 ATATCAAAAAAAGTCCATCTTAATAATTGA 59881

RESULT 35
US-09-949-002-837
; Sequence 837, Application US/09949002
; Patent No. 6900016
; ORGANISM: Human
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 837
; LENGTH: 54286
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-837

Query Match
Best Local Similarity 10.6%; Score 211.4; DB 3; Length 54286;
Matches 267; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 375 TATTAAAGAGCTTTGAATGGCCAGGCGAGTACTCTGCTGTAATCCCAACTTT 434
DB 32756 TCTTCAAAAAACCTGTATAGGCTGGGCGCAGTGTCTCATGTGTAATCCCAACTTT 32815
QY 435 GGAAGGCCAAGTGGGCGGATCACTGAGTCAAGAGTTTAAGACAGGCTGGCAACAT 494
DB 32816 GGAAGGCCAAGTGGGCGGATCACTGAGTCAAGAGTTTAAGACAGGCTGGCAACAT 32875
QY 495 GGTGAACCTGTCTCTACTAATAAAGCAAAATTTAGCCAGTGTGGCATGACCTG 554
DB 32876 GGTGAACCTGTCTCTACTAATAAATAAATAAATTAAGCCAGGAGTGGGCTGCTG 32935
```

Qy	555	TAGTCCCACTACTCAGGAGGTTGAGGAGAGAAATGCTTGAACCTAGAGAGTGAGGT	614
Db	32936	TATGTCCTCCAGCTACTCAGGAGGCTGAGGAGGAAATCGTTGAACCCAGAGGGAGAGGT	32995
Qy	615	TGCATGAACCCGAG--ATGTCACCTGCACCTCAGGCTGGCAAGAGCAACTCCATAA	671
Db	32996	TGCATGAGCCCGCATCATCCCACTGCACTCCAGCCTGGCAACGAGCGAGACCACTCT	33055
Qy	672	AGACAACAAAGCTTTGAATTTGTGAATGAGTTGACT	712
Db	33056	AAAAAAAAAAAAACACACACTTTGTATGGGTTTTCCCT	33096

```

RESULT 36
US-09-949-002-642
: Sequence 642: Application US/09949002
: Patent No. 6900016
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
: TITLE OF INVENTION: AND USES THEREOF
: FILE REFERENCE: CI.000790
: CURRENT APPLICATION NUMBER: US/09/949,002
: CURRENT FILING DATE: 2000-01-28
: PRIOR APPLICATION NUMBER: 60/231,401
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 10823
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 642
: LENGTH: 66955
: TYPE: DNA
: ORGANISM: Human
US-09-949-002-642

```

Query Match	10.6%;	Score 211.4;	DB 3;	Length 66955;
Best Local Similarity	78.3%;	Pred. No. 2.3e-42;		
Matches 267;	Conservative	0;	Mismatches 71;	Indels 3;
				Gaps 1

QY	375	TATTAAAGAGTTTGAATGGGCGCAGGCGCAGTAGTCTCGCTGTAAATCCCAACTTT	434
Db	45237	TCTTCAAAAACACTTGTGATGGGCTGGGGCGAGTGCTCAATGCTGTATATCCAGACATT	45356
QY	435	GGGAGGCCAAGTGTGGCGGATCACTTAGGTCAAGAGTTTAAACACAGCCTGGCCAACT	494
Db	45357	GGGAGGCTGAAGCGCGGCGAGTCACTTGAAGTCAGGAGTTTCAACTAGCTTGGCCAACT	45416
QY	495	GGTGAACCCGTGTCATCTAAAAAGCAAAAATTAGCCAGGCGTGGTGACGACCTG	554
Db	45417	GGTAAAAACCTGTCTCTTCTTAATAATACAAAATTAGCCGGAACAGGTGGCGGGTCTCG	45476
QY	555	TAGTCCCACTACTCAGAGGGTTGAGGAGAGGAAATGTCTGAACCTAGAGGTGAGGT	614
Db	45477	TAGTCCCACTACTCAGAGGCTGAGGACGAGAGGAAATGTCTGAACCCAGAGGCAAGGT	45536
QY	615	TGCAGTACCCCGAG--ATGTCACGTGACCTCACCTGGGCAACAGACCAAGTCCATTA	671
Db	45537	TGCATGAGCCCGCATCTATCCCATCTGCACTCGACCTGGCAACAGACGAGACCATCT	45596
QY	672	AGACAACAAAGCTTGAATTGTGTAAATGAGTTTAACTT	712
Db	45597	AAAAAAAAAAAAACAACACATTTGTATGGGTTTTCCCT	45637

RESULT 37  
US-09-949-016-17270/C  
; Sequence 17270, Application US/09949016  
; Patent No.6812338  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307

```

? CURRENT APPLICATION NUMBER: US/09/949,016
? CURRENT FILING DATE: 2000-04-14
? PRIOR APPLICATION NUMBER: 60/241,755
? PRIOR FILING DATE: 2000-10-20
? PRIOR APPLICATION NUMBER: 60/237,768
? PRIOR FILING DATE: 2000-10-03
? PRIOR APPLICATION NUMBER: 60/231,498
? PRIOR FILING DATE: 2000-09-08
? NUMBER OF SEQ ID NOS: 207012
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 17270
? LENGTH: 12579
? TYPE: DNA
? ORGANISM: Human
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)..(12579)
? OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17270

```

Query Match	10.5%;	Score 210.8;	DB 3;	Length 12579;
Best Local Similarity	58.1%;	Pred. No. 1.8e-42;		
Matches 410;	Conservative 0;	Mismatches 292;	Indels 4;	Gaps 2;

QY I TGATTCACCAACCGCTTGGCTCCGCCAAAGTCGGGATTTACAGGCGGTGAGCCACAGCGCTG 60  
 Db 7608 TCATCTCTCCACCTTGGCTCCGCCAAAGTCGGGATTTACAGGCGGTGAGCCACAGCGCTCTG 7549  
 QY 61 GTCGAATGCTCTTATATTTTGAAGAGCAACATGGGCGCTTAAATCTGTCTTCTATTTGAC 120  
 Db 7548 GTCTGAATTAATTTTACTTGGAAAAATTTCTTGAGAAATTTTCACATGTAAATACATATGCA 7489  
 QY 121 AGACTTTGATGAGTCAAAATCCCAATGCTGCACTTACTGAACGGGCTTAAATGACTTGA 180  
 Db 7488 AAACATCCCAAGAGGACAGCAAAATCTCGCCATGAAAGACTTGTGTGATGTTGAAG 7429  
 QY 181 TCTCTCTAGCTCTTTCTGCAATATGTAAGTGAATATGATGCTTTCAAGAGGAA 240  
 Db 7428 AGAAAGCAATGTGTATATGTATGATAAATTTGAAAACAGTCCAGACAAAGTAACG 7369  
 QY 241 TAAACCTATGAAAGTGTGAGGATAGTGTGATGATGAATPAAGATTTCAACAAGTAG 300  
 Db 7368 AGGGAGAGGAAACGCTTTGATGTAATATGAGACCAAGCTCTTCATATACCTCAGAC 7309  
 QY 301 TAGCTGCTATTGGAAGATTTAAGAGTTATTTATCAACTATTTATATAAATTTTAAAAAC 360  
 Db 7308 TCACATATAGGAGACTCTGTGTTTACTTACCGGTTCTTAAACATATATAGTAGATT 7249  
 QY 361 TAATACACTTAATATTTAAAGACTTTGAAAATGGGACGAGCGAGTAGTCTGCTGCT 420  
 Db 7248 AGCAAGCTCTAAAGTCAGATGTTTGAAAAGAGGAGCCAGGCGAGTGGCTCATCTCTAT 7189  
 QY 421 AATCCCAACCTTTGGGAGGCCAGAGTGGCGGATCACCTGAGGTACAGAGTTTAAAGAC 480  
 Db 7188 AATCCAGACTTTGGGAGGCCAGAGCGGCGAGTCACTTGAGGCGAGAGATTTGAGACC 7129  
 QY 481 AACCTGGCCAAAGATGAGTGAACCCGTGCTACTTAAAAACGCAAAATTTAGCCAGGTG 540  
 Db 7128 AACCTGGCCAAATGAGCGAATCACAGTCTTACTATAAATACAAAAATTTAGTGGGCAATG 7069  
 QY 541 GTGGCATGCACTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAGATTTGCTTGAAC 600  
 Db 7068 GTGGCCGGAACCTGTATATCCCACTATATAGGAGGCTGAGGAGAGAAATTTGCTTAAAC 7009  
 QY 601 TAGAGGTGAGGTTGCAGTAAACCCGAGA--TGTCACTGCACCTCAAGC-TGGCAACAG 656  
 Db 7008 CCGGAGGCGAGAGTTGACGTAGCCAAAGTTGTGCACACTGCACCTCAGCGAGGCAAG 6944  
 QY 657 AGCAAGACTTCATTAAGACAAACAAAGCTTGAATTTGTGTAATG 702  
 Db 6948 AGCGCGACTCTGTCCCAAAAAAAAAAAGTAAAGTGTTCGAAAAAG 6903

```

RESULT 38
US-09-949-016-12682/c
; Sequence 12682, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; ORGANISM: Human
US-09-949-016-12682

Query Match          10.5%; Score 210.6; DB 3; Length 63187;
Best Local Similarity 78.6%; Pred. No. 3,5e-42;
Matches 265; Conservative 0; Mismatches 69; Indels 3; Gaps 1;

QY 349 AATTTTAAACTTAATACCTTAATATTATTTAAAGACTTTGAAATGGCCAGGCGCAGTA 408
DB 35113 AACTTAATTAAGTAACACCATGATAAATAATATAATGACACTTTGGCAGGCATGTG 35054
QY 409 GCTCCGCTGTAAATCCCAACACTTTGGGAGGCGAAGTGGGGGATCACTGAGGTG 468
DB 35053 GCTCATGTCTAATATCCAGCACTTTGGAGGCGCAAGTGGGCAGATCACTGAGGTG 34994
QY 469 GAGTTTAAGACCGAGCTGGGCGCAACATGTGAAACCTGTCTCTACTTAAACGCAAAAAT 528
DB 34993 GAGTTCAAGACCATCTGGCGCAACATGTAAATCTGTCTACTACAAATACAAAAT 34934
QY 529 TAGCCAGGTGTGTGGCATGACCTGTATGTCCCACTACTCAGAGGTTGAGGAGAGTA 588
DB 34933 TAGGAGAGGTATATGGCAACACCTGTAAATTTAGCTACTGGGAGCTGAGGCAAGAGA 34874
QY 589 ATTGCTGAACCTAGGAGGTGAGGTGAGTAACCGAGAT---CTCACTGACCTCCAG 645
DB 34873 ATTGCTGAACCTGGAGGGGAGGTGAGGAGGCAAGATACGCACTGCACTCCAG 34814
QY 646 CCTGGCAACGAGCAAGACTCCATTAAGACAACAAA 682
DB 34813 CCTGGCCACGAGCGCAAGACTCCATTCACAAAAAAA 34777

RESULT 39
US-09-949-016-16288/c
; Sequence 16288, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 16288
; LENGTH: 63187
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16288

Query Match      10.5%; Score 210.6; DB 3; Length 63187;
Best Local Similarity 78.6%; Pred. No. 3.5e-42;
Matches 265; Conservative 0; Mismatches 69; Indels 3; Gaps 1;

OY 349 AATTTTAAACTATATACCTTAATTTTAAAGAGCTTTGAAATGGCGCAGCGAGTA 408
Db 35113 AACTTAATTAATATACCACTATGAAAATTAATTAATTAATGACACTTTGGCCAGCATGTG 35054
OY 409 GCTCTGCTGTATATCCCAACACTTTGGAGGCGCAAGTGGCGGATCACTGAGTGC 468
Db 35053 GCTCATGTCTATTAATCCAGCACTTTGGAGGCGCAAGTGGCGGATCACTGAGTGC 34994
OY 469 GAGTTTAAGACCAAGCTTGGCCCAACTGTGAAACCTGTCTTACTTAAAGCAAAAT 528
Db 34993 GAGTTCAAGACCACTTGGCCCAACTGTGAAATCTGTCTTACTTACAAATACAAAAAT 34934
OY 529 TAGCCAGGTGTGGTGCATNGACCTGTATGTCCTCACTACTCAGAGGTTGAGGAGAGAG 588
Db 34933 TAGGCAAGGTGTGATGGCAACAACCTGTATTTTACTTATGAGGAGGCTGAGGCAAGAG 34874
OY 589 ATTGCTTAACCTTAGAGGTGAGGTGCAGTAACTCCAGAT---GTCACTGACTCCAG 645
Db 34873 ATTGCTTAACCTTAGAGGAGGAGGAGGTTGCAAGGAGCAACAGATCAAGCCACTGCTCCAG 34814
OY 646 CCTGGCAACAGACGACAACTCCATTAAGACCAACAAA 682
Db 34813 CCTGGCCACAGACGACGACACTCCATCTCAAAAAAATA 34777

RESULT 40
US-09-949-016-13713/c
; Sequence 13713, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13713
; LENGTH: 177669
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(177669)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13713

Query Match      10.5%; Score 210.6; DB 3; Length 177669;
Best Local Similarity 80.1%; Pred. No. 5e-42;
Matches 273; Conservative 0; Mismatches 64; Indels 4; Gaps 2;

OY 346 TAAATTTTAAACTATATACCTTAATTTTAAAGAGCTTTGAAATGGCGCAGCGCA 405
Db 90517 TCACATCTTGTATCCCACTAATCTTAAGTGTAGATTAATTAATGAGGCGCAGGATG 90458
OY 406 GTAGTCTGCTGTATATCCCAACACTTTGGAGGCGCAAGTGGCGGATCACTGAGGT 465

```

```
Db 90457 GTGGCTCATGCTCTGTAATCCCGACATTTGGAGGCGCAAGCGGGGTGCTCATTTGAGGT 90398
Qy 466 CAGGAGTTTAAGACCCAGCCCTGGCCACATGCTGAACCCCTGCTCTACTATAAAGCCAA 525
Db 90397 CAGGAGTTTGAAGCCAGCTGGCCAACTGGTGAACCCGCTCTACTATAAATACAA 90338
Qy 526 AATTAGCCAGTGTGTGGCATGCACCTGTAGTCCCACTACTCAGAGGTTGAGGAG 585
Db 90337 AATTAGCTGAGCTGTGTGGCAGACACTGTATCCAGTACTAGGAGGCTGAGGCAAG 90278
Qy 586 AAGATTGCTTGAACCTAGAGGTGAGGTTGACGTACCCGAGAT--GTCACTGCATTC 642
Db 90277 GGAATTGCTTGAACCAAGGGGGCGAGGTTGACGTAGCCAGATCGCCGCTGCATTC 90218
Qy 643 CAGCCTTGGCAACAGCAAGACTCCATTAAGACCAACAAA 682
Db 90217 CAGCCTGGGCAACAGAGTGAAGTCTGTCTCGAAAAAAA 90177
```

```
RESULT 41
US-09-949-016-15924
; Sequence 15924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15924
; LENGTH: 42053
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(42053)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15924
```

```
Query Match 10.5%; Score 210.4; DB 3; Length 42053;
Best Local Similarity 80.2%; Pred. No. 3.4e-42;
Matches 260; Conservative 0; Mismatches 61; Indels 3; Gaps 1;

Qy 376 ATTAAGAGCTTTGAAATGGCCAGGCGCAGTAGCTCTGCTGTAATCCCAACTTTG 435
Db 12080 AATGAAGAGTCAGGGAAGGCGCGGCTGTGCTCATGCTGAATCCCAACTGTG 12139
Qy 436 GGAAGGCCAAGGTGGGCGGATCACTGAGGTGAGAGTTTAAGACCACTTGCCAAATG 495
Db 12140 GGAAGGCCAAGGAGGTAAATCACTGAGGTGAGAGTTCAAGTCAACTGGCCAAATG 12199
Qy 496 GTGAACCTGTGCTCTCTAATAAAGCAAAAATTAGCAAGGTGGTGGTCACTGT 555
Db 12200 GTGAACCTGGCTCTCTCAATAAATTAATAATTAGCTGGCGGTGGTGGCACTGT 12259
Qy 556 AGTCCCACTACTCAGAGGTTGAGGAGAGAAATTGCTTGAACCTTAGAGGTGAGATT 615
Db 12260 AGTCCCAAGTCTCAGAGGTTGAGGAGAGAAATCACTTGAACCCAGGAGGAGATT 12319
Qy 616 GCAAGTAACCGAGAT--GTCACTGCACTCAGCCTGGCAACAGACAACTCCATPAA 672
Db 12320 GCAAGTGAAGTGAATTCGCTACTGCACTCAGCCTGGCAACAGAGAGATCCGCTCT 12379
```

```
Qy 673 GACAACAAAAGCTTTGAATTTGTG 696
Db 12380 AAAAAAAAAAATCAGCAGACTG 12403
```

```
RESULT 42
US-09-949-016-13506/c
; Sequence 13506, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13506
; LENGTH: 283538
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(283538)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13506
```

```
Query Match 10.5%; Score 210.2; DB 3; Length 283538;
Best Local Similarity 77.8%; Pred. No. 7.4e-42;
Matches 267; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

Qy 361 TAATACCTTAATTAATTAAGAGCTTTGAATGGCCAGGCGCAGTAGCTCTGCTGT 420
Db 137661 TCAGGCCACATATTTTAATAAATAGCAAAATGGGCCAGGCAATAGCTCAACTGT 137602
Qy 421 AATCCCAACATTTGGAGGCCAAGGTGGCGGATCACTGAGTCAAGATTTAAGACC 480
Db 137601 AATCCCAACATTTGGAGGCCAAGGTGGCGGATCACTGAGTCAAGATTTAAGACC 137542
Qy 481 AGCTGGCCAAACATGTGTGAACCCCTGCTCTACTATAAAGCAAAAATTAGCAGGTGTG 540
Db 137541 AGCTGACCAACATGTGTGAACCCCTGCTCTACTATAAATTAAGCAGGTGTG 137482
Qy 541 GTGCAATGACCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATTGCTTGAACC 600
Db 137481 GTGCAATGACATATATATCCCACTACTCAGAGGTTGAGGAGAGAAATCGTTGAACC 137422
Qy 601 TAGAGGTGAGGTTGCACTAACCCGAGA--TGTCACTGCACTCCAGCTTGCAACAGA 657
Db 137421 CTGAGAGGTGAGGTTGAGTGAAGTGAAGTGTGCACTGCACTCCAGCTTGCAACAGA 137362
Qy 658 GCAAGACTCCATTAAGACCAAAAAGCTTTGAATTTGTGTA 700
Db 137361 GCAAGACTCCGCTCAATCAATCATCAATAAATGAGATAGA 137319
```

```
RESULT 43
US-09-949-016-16790/c
; Sequence 16790, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
```



;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 16790  
;; LENGTH: 78630  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-16790

Query Match 10.5%; Score 210; DB 3; Length 78630;  
Best Local Similarity 83.0%; Pred. No. 5.3e-42;  
Matches 264; Conservative 0; Mismatches 50; Indels 4; Gaps 2;

QY 356 AAAACATAATACATTAAATTAATTAAGAGCTTGAATGGCCAGGCGAGTAGCTCTCTG 415  
DB 60827 AACTCAAGGAATTAAGAACTTAATTAATGAGGCTCAGGCGCGGTGGCTCAG 60768  
QY 416 CCTGTAATCCCAACACTTTGGAGGCGCAAGGTGGCGGATCACCTGAGTCAAGAGTTTA 475  
DB 60767 CCTATATCCCAAGCACTTTGGAGGCGCAAGGTGGCGGATCACCTGAGTCAAGAGTTTA 60708  
QY 476 AAGACGCTGGCCACATGATGTGAACCCCTGTCTCTACTTAATAAAGCAAAATTAAGCCAG 535  
DB 60707 AAGAACGCTGTCTCAACATGTGTGAACCCCTGTCTCTACTTAATAAATTAAGCCAG 60648  
QY 536 GTGTGTGGCATCAGCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTT 595  
DB 60647 GTGTGTGGCATCAGCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTT 60588  
QY 596 GAACCTAGAGGTGAGGTTGAGTGAACCCGAGA---TGTCACTGCACTTCAGCCTTGCA 652  
DB 60587 GAACCTGGAGGTGAGGTTGAGTGAACCCGAGA---TGTCACTGCACTTCAGCCTTGCA 60528  
QY 653 -ACAGAGCAAGACTCCAT 669  
DB 60527 GACAGAGCAAGACTCCAT 60510

RESULT 44  
US-09-949-016-12396/c  
;; Sequence 12396, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 12396  
;; LENGTH: 78846  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-12396

Query Match 10.5%; Score 209.4; DB 3; Length 78846;  
Best Local Similarity 61.4%; Pred. No. 7.5e-42;  
Matches 422; Conservative 0; Mismatches 256; Indels 9; Gaps 5;

QY 1 TGATCCAGCAGCCTTGCCCTCCCAAGTGTGGATTACAGGCGTAGCCACGACCTG 60  
DB 37780 TGATCCAGCAGCCTTGCCCTCCCAAGTGTGGATTACAGGCGTAGCCACGACCTG 37721  
QY 61 GTGGAATGCTTATTATTATTTGAAGAGCAACATGGCCTTAAATCTGTCTTCAATTGAC 120  
DB 37720 GCCTCCAGTAATGCTTATTATTATTCGACAGAGAAATTTTAAATCCATGCAATTTTCAGGT 37661  
QY 121 AGACTTTGATGAGTGAATATCCCAATGCTGCCACTTACTGAACGGCCTTAAATGACTTG 180  
DB 37660 TTCAATTAATTAATTTATTGACCAAAAGCTATTTCACAAAGCAAGTTCACAAAGTCAAA 37601  
QY 181 TCTCTCAGCTGTCTTCTTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
DB 37600 AAGTACCAATATATCAGTAATCTAATAATGCTGGG--AATCATTTCTTGACAGGATCA 37543  
QY 241 TAAACCTATGAAGAGTGTGAGGATATGTTGATATGAAATTAAGATTTCAACAGTAC 300  
DB 37542 TGATGCTATGGAATATGATGTTTAAATATGATTCAGTTACAAAGCAGTTACAAAGGTTA 37483  
QY 301 TAGCTGCTATGGAATTTAAGATTTATTAAGATTTATTAACAATTTAATTAATTTAATAAC 360  
DB 37482 TCTTATGCAAGCTGCTTGAAGGCAACACCACTGCAAGAAATATGCTCAGGCC 37423  
QY 361 TATACACTTAATTTAATTAAGAGCTTTGAAGTGGCGCAGGCGCAGTACTCTGCTGT 420  
DB 37422 CTACCCAAAGAGATGTTAATAAGAGCAAAAGAGCCGGGTGTGCTCATGCTGT 37363  
QY 421 AATCCCAACTTTGGAGGCGCAAGGTGGCGGATCACTGAGTCAAGAGTTAAGAC 480  
DB 37362 AATCCCAACTTTGGAGGCGCAAGGTGGCGGATCACTGAGTCAAGAGTTAAGAC 37305  
QY 481 AGCTGGCCAAATGTTGAACCTGTCTCTACTTAATAAAGC--AAAAATTAAGCAGTGT 539  
DB 37304 ATCTGTGCTTAACAGGTGAACCCGCTCTCTAATAAATAAATAAATAAATAAATAAATAAATA 37245  
QY 540 GTTGCATCAGCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 599  
DB 37244 GTTGCATCAGCTGTAGTCCCACTACTCAGAGGTTGAGGAGAGAAATGCTTGAAC 37185  
QY 600 CTAGAGGTGAGGTTGAGTGAACCCGAG---ATGTCACTGCACTTCAGCCT--GGCAACA 655  
DB 37184 CCGGAGGCGAGAGCTTGCAGTGAGCGGAGATCATGCACTTCAGCCTTCAGCCTGGCAACA 37125  
QY 656 GAGCAAGATCTCATTAAGACAAACAAA 682  
DB 37124 GAGCAAGATCTCATTAAGACAAACAAA 37098

RESULT 45  
US-09-949-016-12791/c  
;; Sequence 12791, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 12791  
;; LENGTH: 78846  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-12791





```

96 294.2 14.7 915 2 BE729640
97 294 14.7 675 3 BM723230
98 294 14.7 786 5 BU561203
99 293.6 14.7 583 9 AO603128
100 293.6 14.7 713 2 BG773035

```

## ALIGNMENTS

```

RESULT 1
AG15127 717 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-122E08.F, genomic survey sequence.
ACCESSION AG15127
VERSION AG15127.1 GI:16735646
KEYWORDS
SOURCE Pan troglodytes (chimpanzee)
ORGANISM

```

```

Pan troglodytes
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Pan.

```

```

REFERENCE
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Toto, Y., Watanabe, H., and Sakaki, Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 717)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Toto, Y., Watanabe, H., and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center 0045, Japan
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpansec@riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 717
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-122E08.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

```

## COMMENT

```

Query Match 29.6%; Score 592.4; DB 10; Length 717;
Best Local Similarity 97.1%; Pred. No. 1.1e-85;
Matches 635; Conservative 0; Mismatches 16; Indels 3; Gaps 3;

```

## FEATURES

## source

```

1. 717
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-122E08.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

```

## ORIGIN

```

Query Match 29.6%; Score 592.4; DB 10; Length 717;
Best Local Similarity 97.1%; Pred. No. 1.1e-85;
Matches 635; Conservative 0; Mismatches 16; Indels 3; Gaps 3;

```

```

DB 243 GGAGGGGGGATTTGAAGGTGAAGCCAGCTGGCTTCTGGCTCAGGTGGGACTTGTGAAGA 302
451 CTTTGTGTGCTAGCTAAAGATTGTAAATGACCAATGAGCACTGTGTCTAGCTAAAG 510
303 CTTTGTGTGCTAGCTAAAGATTGTAAATGACCAATGAGCACTGTGTCTAGCTAAAG 362

```

```

DB 511 GATTGTAATGACCAATCAGCACTCTGTAAATGACCAATCAGCAAGATGTGGGCGG 570
363 GATTGTAATGACCAATCAGCACTCTGTAAATGACCAATCAGCAAGATGTGGGCGG 422

```

```

DB 571 GTCAATTAAGGAGTAAATCTGGCCACCCAGCCAGCAAGTGGCAACCACTGGGCTCC 630
423 GTCAATTAAGGAGTAAATCTGGCCACCCAGCCAGCAAGTGGCAACCACTGGGCTCC 482

```

```

DB 631 CTTCCACCTGTGGAAGCTTTGTTCTTTGCTCTTCAATTAATCTTGTCTGCTCAT 690
483 CTTCCACCTGTGGAAGCTTTGTTCTTTGCTCTTCAATTAATCTTGTCTGCTCAT 542

```

```

DB 691 TCTTTGTGTCACACTA-CCTTTATGAGCTGTAACTCACTGCGAGGCTGTGGCTTC 749
543 TCTTTGTGTCACACTA-CCTTTATGAGCTGTAACTCACTGCGAGGCTGTGGCTTC 602

```

```

DB 750 ATTCTGAAGTCAACAGACCAAGAA-CCCACTGGAAGAACAA-GAATCCCATGTGC 807
603 ATTCTGAAGTCAACAGACCAAGAA-CCCACTGGAAGAACAA-GAATCCCATGTGC 662

```

```

DB 808 TGCCTTTAAGACTGTAACTCACTGGAAGCTTGCAGCTTCACTCTGTAAG 861
663 CCGCTTAAGACTGTAACTCACTGGAAGCTTGCAGCTTCACTCTGTAAG 716

```

```

RESULT 2
BM041081 855 bp mRNA linear EST 07-NOV-2001
LOCUS BM041081
DEFINITION 603614811p1 NIH_MGC_108 Homo sapiens cDNA clone IMAGE:5556482 5',
mRNA sequence.
ACCESSION BM041081
VERSION BM041081.1 GI:16770336
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM

```

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 855)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgrabbs-rc@mail.nih.gov
Tissue Procurement: ATCC

```

```

REFERENCE
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgrabbs-rc@mail.nih.gov
Tissue Procurement: ATCC

```

```

cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHC82020 row: f column: 03
High quality sequence stop: 819.
Location/Qualifiers
1. 855
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5556482"
/tissue_type="M1ms" tumor, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 108"
/notes="Organ: Kidney; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACAG(G). Library constructed by Ling Hong in the

```

## FEATURES

## source

## ORIGIN

Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

Query Match 25.2%; Score 504.6; DB 3; Length 855;  
Best Local Similarity 84.8%; Pred. No. 1.5e-71;  
Matches 623; Conservative 0; Mismatches 84; Indels 28; Gaps 4;

```

QY 334 AGGAAGTACGCTACAGCGGTGATC-GCCAAATCCCAACAGACGTGGGGTCTCTGTTTGG 392
DB 2 AGGAAGTACGCTACAGCGGTGATCACCACCAATCCCAACAGACGTGGGGTCTCTGTTTAA 61
QY 393 AGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTACGTGGGACCTTGAAGACT 452
DB 62 TGGGGAGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTGGGTGGGACCTTGAAGACT 121
QY 453 TTGTGCTAGCTAAAGGATGTAAATGACCAATCAGACCTGCTGTCTAGCTAAAGGA 512
DB 122 TTCTGTCTAGCTAAGAGATTGTAACACCAATCAGCTAGCTAAGTGTCTAGCTAAGAGT 181
QY 513 TTGTAATGTC-----ACCAATCAGACCTCTGTAAATGGA 547
DB 182 TTGTAATGACCAATCAGACCTCTGTAAATAACGACCAATCAGACCTCTGTAAATGGA 241
QY 548 CCAATCAGACGATGTGGGGGGGTCAATTAAGGAGTAAATCTGGCCACCCGACCG 607
DB 242 CCAATCAGAGATGTGGGGGGGCAAAATTAAGGAGTAAATCTGGCCACCTGAGTCTAG 301
QY 608 CAGTGGCAACCCAGCTGGGGGCCCTTCCACAGTGGAGAGCTTGTCTTTTGGCTCTGA 667
DB 302 CAGTGGCAACCCAGCTGGGGGCCCTTCCACAGTGGAGAGCTTGTCTTTTCACTCTGA 361
QY 668 CAATAATCTGTGCTGCTCATCTTGTGTGTCACACTACTTATGAGCTGTAACT 727
DB 362 CAATAATCTGTGCTGCTCATCTTGTGTGTCACACTACTTATGAGCTGTAACT 421
QY 728 CACTGCGAGGGTCTGTGGCTTCACTTCTGTAAGTCAAC-AGACCAAGAACCCACTGGAAG 786
DB 422 CACTGCGAGGGTCTGTGGCTTCACTTCTGTAAGTCAACAGACCAAGAACCCACTGGAAG 481
QY 787 AACAAAGAACTCCGATGTGTCGCTTTAAGAGCTGTAACTCATCTGAGAGCTGTGA 846
DB 482 AACAAAGAACTCCGATGTGTCGCTTTAAGAGCTGTAACTCATCTGAGAGCTGTGA 541
QY 847 GCTTCACTCTGTAAGTCAAGTGAAGCAAAACCCACAGAGGAAATCTGTGACAG 906
DB 542 GCTTCACTCTGTAAGTCAAGTGAAGCAAAACCCACAGAGGAAATCTGTGACAG 601
QY 907 CCTGAATATCTGAAGAACTCAAGACCACTCTTTCAGAGCTGTAACTGAC 966
DB 602 TCGAACATCTGTAATGAACAACTGTGACACCGCACTTTTAAGAACTGTAACTGAC 661
QY 967 GCAAGGGTCTGTGGCTTCAATCTTG-AGTCAAGAGACCAAGAACCCGGAAGAGAC 1025
DB 662 GTGAGGGTCTGTGGCTTCAATCTTG-AGTCAAGAGACCAAGAACCCGGAAGAGAC 721
QY 1026 AAATTCCAGACAG 1040
DB 722 CAATTCCGACACAG 736

```

RESULT 3  
B2608243/c  
LOCUS B2608243 836 bp DNA linear GSS 08-JUN-2003  
DEFINITION WHACT79TR Human MCF7 breast cancer cell line library (MCF7.1) Homo  
sapiens genomic clone MCF7\_1-17N14, genomic survey sequence.  
ACCESSION B2608243  
VERSION B2608243.1 GI:31516804  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBLISHED  
COMMENT

1 (bases 1 to 836)  
Volik, S., Zhao, S., Chin, K., Brebner, J. H., Herndon, D. R., Tao, Q.,  
Kowbel, D., Huang, G., Lapuk, A., Kuo, W.-L., Magrane, G., de Jong, P.,  
Gray, J. W., and Collins, C.  
End-sequence profiling: Sequence-based analysis of aberrant genomes  
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)  
12788976  
Contact: Volik SV  
Colin Collins' lab  
UCSF Comprehensive Cancer Center  
UCSF Box 0808, San Francisco, CA 94143-0808, USA  
Tel: 415 502 7066  
Fax: 415 502 5665  
Email: svolik@cc.ucsf.edu  
This clone is available from Amplicon Express  
http://www.genomex.com  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..836  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="MCF7\_1-17N14"  
/sex="female"  
/clone\_1ib="Human MCF7 breast cancer cell line library (MCF7.1)"  
/note="Vector: pECBAC1; Site 1: HindIII; This library was constructed from MCF7 breast cancer cell line by Amplicon Express (http://www.genomex.com) using their standard procedure."

## ORIGIN

Query Match 23.8%; Score 476.8; DB 9; Length 836;  
Best Local Similarity 87.6%; Pred. No. 4.6e-67;  
Matches 589; Conservative 0; Mismatches 52; Indels 31; Gaps 5;

```

QY 1 AAAGGCTTGAATACAGCAATGCTTCAACTCTTATACCACTCTGAGTTGGG 60
DB 669 AAAGGCTTGAATACAGCAATGCTTCAACTCTTATACCACTCTGAGTTGGG 610
QY 61 GACATGGCTTCTCCCTTTCTAGGTCCTGTGACAGCATCTTGTATATGTGCAATTGG 120
DB 609 GACATGGCTTCTCCCTTTCTAGGTCCTGTGACAGCATCTTGTATATGTGCAATTGG 550
QY 121 GCCGTGATTTTAACTCTTGTGCAAAATTTGTTCTCTAGAGATGAGGCCATCAAGT 180
DB 549 GCCGTGATTTTAACTCTTGTGCAAAATTTGTTCTCTAGAGATGAGGCCATCAAGT 490
QY 181 ACAGATGATCTTACAAATGTAAACCCCAATGAGCTCAACTCAACTTCTGTGAGAGCC 240
DB 489 ACAGATGATCTTACAAATGTAAACCCCAATGAGCTCAACT---AACTTCTAGAGAGCC 433
QY 241 CCTGAGACGACCGGCTGGGCTTTCAATGGCTTAAAGAGCTCCCTCTGAGAGACATAC 300
DB 432 CCTGAGACGACCGGCTGGGCTTTCAATGGCTTAAAGAGCTCCCTCTGAGAGACATAC 373
QY 301 CACTGAGGGGCCCTTCTTCAACCCCTATCAGACGAAAGTAGCTACAGGGTCTATC-GGC 359
DB 372 AACTGAGGGACCTTCTTGGCCCTATCAGACGAAAGTAGCTACAGGGTCTATCAGCC 313
QY 360 AAATCCCAACAGACCTGGAGTGTCTGTTTGAAGGGGAGTTGAAGGTGAAGCAAGCT 419
DB 312 AAATCCCAAGTACAGTGGGGTGTCCGTTTGAAGGGGAGTTGAAGGTGAAGCAAGCT 253
QY 420 GGGCTT-CTGGGTACAGTGGGAGCTTGAAGAACTTTGTGTACTATAAGATTTGTA 478
DB 252 GGAATTTCTGATCAAGTGGGAGCTTGAAGAACTTTGTGTACTATAAGATTTGTA 193
QY 479 TGCACCAATCAGCA-----CTCTGTGTCTAGCTAAAGAT 513

```

Db 192 TGCACCAATCAGCACTCTATATAAATGACCAATCAGCGCTGTGTCTAGCTAAAGAT 133  
QY 514 TGTAAATGACCAATCAGCACTCTGTAAATGAGCAATCAGAGATGCGGCGGATC 573  
Db 132 TGTAAACACACCAATCAGCACTCTGTAAATGAGCAATCAGAGATGCGGCGGATC 73  
QY 574 AATAAGGAGATTAATAAATGAGCAATCAGAGATGCGGCGGATC 632  
Db 72 AGTAAGGAGATTAATAAATGAGCAATCAGAGATGCGGCGGATC 13  
QY 633 TCCACACTGTGG 644  
Db 12 TCCACACTGTGG 1

RESULT 4  
LOCUS AG045844 655 bp DNA linear GSS 02-NOV-2001  
DEFINITION Pan troglodytes DNA, clone: PTB-024004.R, genomic survey sequence.  
ACCESSION AG045844  
VERSION AG045844.1 GI:16582736  
KEYWORDS GSS  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.

REFERENCE 1  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 655)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenhiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.  
PRIMERS  
Sequencing: MJ3Rev  
LIBRARY  
Vector : PKS145  
R.site 1 : SacI  
R.site 2 : SacI.  
Location/Qualifiers  
1. 655  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-024004.R"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"

ORIGIN  
Query Match 22.9%; Score 458; DB 10; Length 655;  
Best Local Similarity 89.2%; Pred. No. 5,4e-64;  
Matches 505; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

QY 489 AGCACTCTGTGTCTAGCTAAAGATTTGTAATGACCAATCAGCACTCTGTAAATGAGC 548  
Db 50 AGAGCTCTGTATCTAGCTAAAGATTTTATATATACCAATCAGCACTCTGTAAATGAGC 109  
QY 549 CAATCAGCAGATGTGGCGGGGCAATAAGAGTAAGTAAGTGGCCAGCCGACGAGC 608  
Db 110 CAATCAGCAGATGTGGCGGGGCAATAAGAGTAAGTAAGTGGCCAGCCGAGC 169

QY 609 AGTGGCAACCACTGGGGTCCCTTCCACACTGTGAAAGCTTTGTCTTTGCTCTTAC 668  
Db 170 AGCGGCAACCACTGAGGTCCTTCCACTCTGTGAAAGCTTTGTCTTTGCTCTTAC 229  
QY 669 AATAATCTTGTGTCTCATTTCTTTGTGTCACACTATCTTTATAGCTGTACATC 728  
Db 230 AATAAATCTTGTGTCTCATTTCTTTGTGTCACACTATCTTTATAGCTGTACATC 289  
QY 729 ACTGGAGGGTGTGGCTTCAATTCCTGAAGTCAAC-AGACAGCAACCACTGAGAGA 787  
Db 290 ACCAGAGGGTGTGGCTTCAATTCCTGAAGTCAACAGCAACCACTGAGAGA 349  
QY 788 ACAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCTACTGGAAGCTGTGAG 847  
Db 350 ACAAGAACTCTGATGTGCTGCTTTAAGAGCTGTAACTCTACTGGAAGCTGTGAG 409  
QY 848 CTTCACCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 907  
Db 410 CTTCACCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 469  
QY 908 CTGAATATCTGAAGGAACCACTCCAGACACACATCTTTAGAGCTGTAACTCACTCAC 967  
Db 470 CTGAATATCTGAAGGAACCACTCCAGACACACATCTTTAGAGCTGTAACTCACTCAC 529  
QY 968 CAAGGCTGTGGCTTCAATTTGTAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1027  
Db 530 CAAGGCTGTGGCTTCAATTTGTAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAG 589  
QY 1028 ATTCCAGACACAGTGAATCTGTA 1053  
Db 590 ATTCCAGACACAGTGTGTGACTGCA 615

RESULT 5  
LOCUS CR745511 672 bp mRNA linear EST 21-DEC-2004  
DEFINITION CR745511 Soares testis NMT Homo sapiens cDNA clone IMAGE:757500  
ACCESSION CR745511  
VERSION CR745511.1 GI:51667764  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 672)  
AUTHORS Ebert, U., Heil, O., Hennig, S., Korn, B., Neubert, P., Partsch, R., Peters, M., Radelof, U. and Schneider, D.  
TITLE I.M.A.G.E. cDNA Clone Collection  
JOURNAL Unpublished (2004)  
COMMENT Contact: Inge Airlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Email: www.rzpd.de  
RZPD: IMAGP98P131862.  
RZPDLIB: I.M.A.G.E. cDNA Clone Collection;  
Contact: Inge Airlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 100  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available royalty-free from RZPD;  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
M13r. Primer sequence: TTTCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1. 672  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGP98P131862 ; IMAGE:757500"  
/sex="male"

## ORIGIN

Query Match 22.8%; Score 456.6; DB 7; Length 672;  
Best Local Similarity 89.3%; Pred. No. 8.9e-64;  
Matches 503; Conservative 0; Mismatches 59; Indels 1; Gaps 1;

/lab host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5].  
TGTTCACATCTGAGTGGGAGCGCGCCATTTTTTTTTTTT 3'.  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to Cos5, and was  
constructed by Bento Soares and M. Fatima Ronaldo.

485 AATCAGCACTCTGTCTAGCTAAGATGTAAATGACCATTCGACTCTGTAAAT 544  
1 AATCAACACTCTGTCTGTAGTAAGATGTAAATGACCATTCGACTCTGTAAAT 60  
545 GGAACCAATCAGCAGATGTGGCGGGTCAATTAAGGAGTAAAACTGGCCACCGAGC 604  
61 GGAACCAATCAGCAGATGTGGCGGGTCAATTAAGGAGTAAAACTGGCCACCGAGC 120  
605 CAGCAGTGCACACCTCGGGTCCCTTCACACTGTGAGACTTTTCTTCTGCTCT 664  
121 CAGCAGTGCACACCTCGGGTCCCTTCACACTGTGAGACTTTTCTTCTGCTCT 180  
665 TCACATTAATCTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 724  
181 TCACATTAATCTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 240  
725 ACTCAGTGCAGAGGCTGTGGCTTCAATCTGTAAGTCAAC-AGACCAAGAACCACTGTA 783  
241 ACTCAGTGCAGAGGCTGTGGCTTCAATCTGTAAGTCAAC-AGACCAAGAACCACTGTA 300  
784 AGGACCAAGAACCTCCGATGTGCTGCTCTGTAAGTCAAC-AGACCAAGAACCACTGTA 843  
301 AGGACCAAGAACCTCCGATGTGCTGCTCTGTAAGTCAAC-AGACCAAGAACCACTGTA 360  
844 GCAAGCTTCACTCTGTAAGTCAAC-AGACCAAGAACCACTGTAAGTCAAC-AGACCAAG 903  
361 GCAAGCTTCACTCTGTAAGTCAAC-AGACCAAGAACCACTGTAAGTCAAC-AGACCAAG 420  
904 ACACCTGAATATCTGTAAGTCAAC-AGACCAAGAACCACTGTAAGTCAAC-AGACCAAG 963  
421 ACACCTGAATATCTGTAAGTCAAC-AGACCAAGAACCACTGTAAGTCAAC-AGACCAAG 480  
964 ACCGAGAGGCTGTGGCTTCAATCTGTAAGTCAAC-AGACCAAGAACCACTGTAAGTCA 1023  
481 ACCGAGAGGCTGTGGCTTCAATCTGTAAGTCAAC-AGACCAAGAACCACTGTAAGTCA 540  
1024 ACAATTTCCAGACAGTGAAG 1046  
541 ATTAATTCAGACACCAAGCA 563

RESULT 6  
LOCUS A0745404 860 bp DNA linear GSS 16-Jul-1999  
DEFINITION HS\_2276\_A1\_G08\_T7C CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=2276 Col=15 Row=M, genomic survey  
sequence.

ACCESSION A0745404  
VERSION A0745404  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

## REFERENCE

1 (bases 1 to 860)  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, W.D., and  
Hood, L.

## TITLE

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

## JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

## COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 2276 row: M column: 15

Seq primer: T7

Class: BAC ends

High quality sequence stop: 860.

## FEATURES

Location/Qualifiers

1..860

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="Plate=2276 Col=15 Row=M"

/sex="male"

/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in  
E-Coli DH10B"

## ORIGIN

Query Match 22.5%; Score 450.8; DB 9; Length 860;  
Best Local Similarity 87.8%; Pred. No. 7.2e-63;  
Matches 526; Conservative 0; Mismatches 68; Indels 5; Gaps 3;

1 AAAGCTTTGTAATGACAAATGCTTTTCAAACTTTATACCACTCTGAGTTGGC 60  
701 AAAGCTTTGTAATGACAAATGCTTTTCAAACTTTATACCACTCTGAGTTGGC 645  
61 GACATGGCTTCCCTCTCTAGGTCGTCGACAGCATCTGTAATGTCATTTGG 120  
644 AACATGGCTTCTACCTTTCTAGGTCGTCGACAGCATCTGTAATGTCATTTGG 585  
121 GGCCTGATTTTAACTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 180  
584 GGCCTGATTTTAACTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 525  
181 ACAGATGATCTTAACTGTAATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 240  
524 ACAGATGATCTTAACTGTAATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 465  
241 CCTGACGAGACCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 300  
464 CCTGACGAGACCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 405  
301 CACTGACGAGACCGGCTTCACTCCCTATCCAGCAGGAATGACTACAGGTCATC-GCC 359  
404 AACTGACGAGACCGGCTTCACTCCCTATCCAGCAGGAATGACTACAGGTCATCACC 345  
360 AAATCCCAACAGAGCTGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 419  
344 AAATCCCAACAGAGCTGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 285  
420 GGGCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 479  
284 GGGCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 225  
480 GCAACCAATCAGACTCTGTCTAGCTAAAGATGTAAATGACCAATCAGACTCTGT 539  
224 GCATCAATCATTTGCTGTCTAGCTAAAGATGTAAATGACCAATCAGACTCTGT 165



QY 540 -AAATGACCAATGACAGATGCGGGGTCAATAATGAGAGTAAATCTGGCCA 597  
164 AAAAAAGCAACCAATGACCACTAATGAAATGACCAATGACCACTTTAAATGACCA 106

RESULT 7  
AGI69322 680 bp DNA linear GSS 09-JAN-2002  
LOCUS Pan troglodytes DNA, clone: RP43-037P09.TU, genomic survey  
DEFINITION sequence.  
ACCESSION AGI69322.1 GI:16699000  
VERSION AGI69322.1  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Pan.

REFERENCE  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of library RPCI-43  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 680)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Ageo Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Shuhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimbases@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC  
end was generated during the Rad process and may have higher chance  
of clone tracking errors.  
PRIMERS  
Sequencing: TU  
LIBRARY  
Vector : pBAC3.6  
R.Site 1 : EcoRI.  
R.Site 2 : EcoRI.  
Location/Qualifiers  
1. 680  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="RP43-037P09.TU"  
/sex="male"  
/cell\_type="lymphocytes"  
/clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

ORIGIN  
Query Match 22.3%; Score 445.8; DB 10; Length 680;  
Best Local Similarity 88.0%; Pred. No. 4.9e-62;  
Matches 530; Conservative 0; Mismatches 68; Indels 4; Gaps 4;

QY 400 ATTGAGAGTGAGCCAGCTGGGCTCTGGTCAAGTGGGAGCACTTGGAACCTTTGTG 459  
80 ACTGAGGGGTGAAGCCAGCTGGGTTCTGGGTGGGGAGTGGAGAACTTTCTGT 139

QY 460 CTAGCTAAAGATTTGTAATGACCAATGACCACTGTGTCTAGCTAAAGATTGTA 519  
140 CTAGCTAAAGATTTGTAATGACCAATGACCACTGTGTCTAGCTAAAGATTGTA 199

QY 520 TGACCAATGACCACTGTGT-AAAATGACCAATGACCAAGATGGGCGGGCTCAATA 578  
200 CACACCAATGACCACTGTGTAAATGACCAATGACCAATGATGGGCGGCAATA 259

QY 579 AGGAGTAAATGACCTGGGCAAGCCAGCAAGATGGGCAAGCCATCGGGTCCCTTCACA 638  
260 AGGGAATAAAGCTGGCCACCC-AGCTAAGAGGGCCACCACTAAGGATACCTTCATG 318

QY 639 CTGCGAAGCTTTGTTCTTTTCTCTTCAATTAATCTTGCTGCTCATTTGTTG 698  
319 CTGCGAAGCGTTGTTCTTTTCTCTTCAATTAATCAATGCTGCTCATTTGTTG 378

QY 699 TCCACACTACTTTTGTAGCTGTAACTCACTGAGAGGCTGTGGCTTCACTTCTGA 758  
379 TCTGACACTCTTTATAGCTGTAACTTTGTGGAGGTGTGGCTTCACTTCTGA 438

QY 759 GTCA-ACAGACCAGCAACCCACTGAGAGCAAGAACTCCCGATGTGCTCTTAA 817  
439 GTGAGAGAAACCAAGAACCCAGAGAGGAATGAACAACCTCCAGAGCACTTAA 498

QY 818 AGCTGTAACTACTGCGAGCTTGAAGCTTCACTCTCTTGAAGTGTAGACACAA 877  
499 AGCTGTAACTACTGCGAGCTTGAAGCTTCACTCTCTTGAAGTGTAGACACCA 558

QY 878 CCCACAGAGGAAGAAACTGTGACACACCTGAATATCTGAAGAAACAACTCCAGA 937  
559 CCCACAGAGGAAGAAACTGTGACACATATCTGAACATGAGAGAAACAACTTGA 618

QY 938 CACCACTCTTCAAGCTGT-TAACTTCAACCGAAGGCTGTGCTTCACTTCTGAAGTC 996  
619 CACCACTTGAAGCTGTAACTTCACTTCACTTCACTTCACTTCACTTCACTTGAAGTC 678

QY 997 AG 998  
679 AG 680

RESULT 8  
BX507549/c 649 bp mRNA linear EST 04-SEP-2003  
LOCUS DKFP68611625 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
DEFINITION DKFP68611625 5', mRNA sequence.  
ACCESSION BX507549  
VERSION BX507549.1 GI:32040300  
KEYWORDS EST.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE  
AUTHORS Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R.,  
Mewes, H. W., Well, B., Amid, C., Osanger, A., Fobo, G., Han, M. and  
Wiemann, S.  
TITLE EST (Poustka, A., Albert, R., Moosmayer, P., Schupp, I.,  
Wellenreuther, R., et al.)  
JOURNAL Unpublished (2003)  
COMMENT Contact: MIPS  
MIPS  
Ingolstaedter Landstr.1, D-85764 Neuberg, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@ktz-heidelberg.de;  
sequenced by DKFZ (German Cancer Research Center,  
Heidelberg/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
al sequence also available.  
This clone (DKFP68611625) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1. 649  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFP68611625"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="686 (synonym: hlcc3)"  
/note="Vector: pTriplex2; Site\_1: SfiI; Site\_2: SfiI;  
cDNA-collection"

## ORIGIN

Query Match 22.0%; Score 440.8; DB 5; Length 649;

Best Local Similarity 88.9%; Pred. No. 3,2e-61; Matches 514; Conservative 0; Mismatches 52; Indels 12; Gaps 3;

476 AAATGACCAATAGCACTGTGTCTAGCTTAAAGATTGAAATGACCAATCAGACT 535  
 610 AAATGACCAATAGCACTGTGTCTAGCTTAAAGATTGAAATGACCAATCAGACT 551  
 536 CTGT-AAAATGACCAATCAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 594  
 550 CTGTAAAAATGACCAATCAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 491  
 595 CCACCCGAGCCAG 654  
 490 CCACCCGAGCCAG 431  
 655 CTTTGTCTCTTCAAAATTAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 714  
 430 CTTTGTCTCTTCAAAATTAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 371  
 715 GAGCTGTACACTGCTGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 773  
 370 GAGCTGTACACTGCTGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 311  
 774 ACCCACTGAAAGAAACAAAGATCTCCGATGTGTGTGTGTGTGTGTGTGTGT 833  
 310 GCCCAGCAG 251  
 834 GCGAGAGCTGTGAGCTTCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 893  
 250 GCGAGAGCTGTGAGCTTCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 191  
 894 AACTGTGACACACTG-----ATATCTGAAAGAAACAACTCCAGACACCAT 943  
 190 AACTGTGACACACTGAAATGAGACATCTGAAAGAAACAACTCCGAGACACCAT 131  
 944 CTTTCAGAGCTGTACACTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGT 1003  
 130 CTTTAAGAACTGTAACTGACCTGACCGGAGGTGTGTGTGTGTGTGTGTGTGT 71  
 1004 CCAAGAACCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1041  
 70 CCAAGAACCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 33

RESULT 9  
 B72370 621 bp DNA linear GSS 08-Apr-1999  
 LOCUS B72370  
 DEFINITION RPEC111-8U21.TP RPEC111 Homo sapiens genomic clone RPEC111-8U21,  
 genomic survey sequence.  
 ACCESSION B72370  
 VERSION B72370.1 GI:2711521  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 621)  
 Adams,M.D., Rounsley,S.D., Field,C.E., Baas,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
 Venter,J.C.  
 Use of BAC End Sequences for Sequence-Ready Map Building  
 Unpublished (1997)  
 Other GSSs: RPEC111-8U21.TV  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: mdadams@tigr.org  
 Clones are derived from the human BAC library RPEC111. For BAC  
 library availability, please contact Pieter de Jong  
 (pieterdejong.med.bufileo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from  
 Research Genetics (info@resgen.com). BAC end search page:  
 http://www.tigr.org/db/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: SP6  
 Class: BAC ends.

## FEATURES

Location/Qualifiers  
 1..621  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GB:7502924"  
 /clone="RPCT-11-8U21"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /clone\_lib="RPCT-11"  
 /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
 RPEC111 Human Male BAC Library"

## ORIGIN

Query Match 21.5%; Score 429.8; DB 9; Length 621;  
 Best Local Similarity 84.9%; Pred. No. 1.9e-59;  
 Matches 527; Conservative 0; Mismatches 67; Indels 27; Gaps 3;

291 AGGACACTACCACTGAGAGGCCCCCTTCTTACCCCTATCAGAGAGAGTACTACAGC 350  
 1 AGGACACTACCACTGAGAGGCCCCCTTCTTACCCCTATCAGAGAGAGTACTACAGC 60  
 351 GTTCATC-GCCAAATCCCAACAGCAGTGGGGTCTCTGTTTGAAGGGGGAATTGAGAGT 409  
 61 GTTCATGAGCAATTCACACAGCATTTGGGGGTCTCTGTTTGAAGGGGGAATTGAGAGT 120  
 410 GAAGCAGCTGGGCTTCTGGGTCAAGGTGGGAGCTTGAAGAACTTTGTCTAGCTAAAG 469  
 121 GAAGCTGGCTGGGCTTCTGGGTCAAGGTGGGAGCTTGAAGAACTTTGTCTAGCTAAAG 180  
 470 GATTGAAATGACCAATGAGACCTGTGTCTAGCTAAAGATTTGAAATGC----- 522  
 181 GATTGAAATGACCAATGAGACCTGTGTCTAGCTAAAGATTTGAAATGC----- 240  
 523 -----ACCAATGAGCACTCTGTAAATTTGACCAATGAGAGATTTG 564  
 241 AGCACTCTGTAAATTAAGACCAATTAAGCACTCTGTAAATTTGACCAATGAGATTTG 300  
 565 GCGGGGGTCAATTAAGGAGTAATAAAGTGGCCAGCCAGCCAGAGTGGCAACCCAGCTG 624  
 301 GGTGGGGCCAAATTAAGGAGTAATAAAGTGGCCAGCCAGCCAGAGTGGCAACCCAGCTG 360  
 625 GGTCCCTTTCACACTGTGGAAGCTTTGTCTTTTGTCTTTCACATTAATCTTGTCTGT 684  
 361 GGTCTCTTTCACAGCTGTGAAAGCTTTGTCTTTTGTCTTTCACATTAATCTTGTCTGT 420  
 685 GCTCATCTTTTGTGTCAACTACCTTTATGAGCTTTACACTCACTGGAAGGTCTGTG 744  
 421 GCTCATCTTTTGTGTGTCAACTACCTTTATGAGCTTTACACTCACTGGAAGGTCTGTG 480  
 745 GCTTCATCTCTGAAGTCAAC-AGACCAAGAACCCAGCTGGAAGGAACAAAGATCCCGAGT 803  
 481 GCTTCTCTCTGAAGGAG 540  
 804 GTGCTGCTTTTGAAGCTTTGAACACTGAGAGAGCTTGAAGCTTCACTCTGAAGT 863  
 541 ATGCCACCTTTTGAAGCTTTGAACACTGAGAGAGCTTGAAGCTTCACTCTGAAGT 600  
 864 AGTGAGACCAAGAACCCAGCA 884  
 601 AGTGAGACCAAGAACCCAGCA 621

## RESULT 10

B0677568 940 bp mRNA linear EST 15-JUL-2002  
LOCUS B0677568  
DEFINITION AGENCOURT 8498032 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6257721  
5', mRNA Sequence.  
ACCESSION B0677568  
VERSION B0677568.1 GI:21790247  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 940)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabds-remail.nih.gov](mailto:cgabds-remail.nih.gov)  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LICM2412 row: h column: 10  
High quality sequence start: 22  
High quality sequence stop: 675.  
Location/Qualifiers  
1..940  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6257721"  
/rissue\_type="melanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN  
Query Match 21.1%; Score 424.2; DB 5; Length 940;  
Best Local Similarity 84.7%; Pred. No. 1.3e-58;  
Matches 557; Conservative 0; Mismatches 69; Indels 32; Gaps 6;  
QY 154 TTCCTTAGAGTGAAGGAGCCATCAAGCTACAGATGATCTTAACAATGAACCCCAATGAG 213  
DB 263 TTGATTGAGATGAGGCTATCAAGCTACAGATGATCTTAACAATGAACCCCAATGAG 322  
QY 214 CTCACCTAACCACTTCTGTGAGAACCCCTGAGACCGCGCTGGCCCTTTCAATGAGCCT 273  
DB 323 CTCACCTAACCA---CTACCAAGGACCCCTGAGCAACCGCGCTGGCCCTTTCAATGAGCCT 379  
QY 274 AAGAGCTCCCTCTGAGAGACACTACACTGAGGAGCCCTTCTTCAACCCCTATCCAGC 333  
DB 380 AAGAGATTTCCCTCTGAGAGACACTACACTGAGGAGCTCTTTGCCCCCTATCCAGC 439  
QY 334 AGAATAGCTACAGGCGTATC-GCCAATCCCAAGAGCTGGGGTCTCCTGTTTG 392  
DB 440 AGAAGTAGCTACAGTGTATCACCCAATTTCCCAAGAGTGGGGTCTTTGTTAAG 499  
QY 393 AGGGGGATTGAGAGTGAAGCAGCTGGCTTCTGGGTGAGGTGGGACTTGGAGAACT 452  
DB 500 TGGGAGATTGAGAGGTGAAGCAGCTGGCTTCTGGGTGGGTGGGACTTGGAGAACT 559  
QY 453 TTGTGTCTAGCTAAAGATTGTAATGCAACAATCAGCACTGTGTCTAGCTAAAGA 512  
|||||

DB 560 CTTCGTCTAGCTAGAGATTGTAAACACACCAATCAAGTGTATGTCTAGCTAGAGGT 619  
QY 513 TTGTAATATGTC-----ACCAATCAGCACTCTGTAAATGGA 547  
DB 620 TTGTAATATGCAACCAATCAGCACTCTGTAAATGGA 679  
QY 548 CCAATCAGCAGATGTGGGCGGGGTCAATTAAGGAGTAAATAGTGGCCACCGGACGAG 607  
DB 680 CCAATCAGTATGATGCGGGCAGGCGCAATTAAGGAGTAAAGCTGGCCACCTGAGTCAG 739  
QY 608 CAGTGCAAC-CCCACTGGGGTCCCTTCACACTGTGGAAGCTTTGTTGCTTC 666  
DB 740 CAGTGCAACCCCACTGGGGTCCCTTCACACTGTGGAAGCTTTGTTGCTTC 799  
QY 667 ACAATTAATCTTGTCTGTCT-CACTTTTGTGTCACTACTTATGAGCTGTAAACA 725  
DB 800 ACAATTAATCTTGTCTGTCTCTCACTCTTTGGGTCAACCACTTATGAGCTCAACC 859  
QY 726 CTCACCTGCGAGGCTGTGGCTTCACTTCTGTAAGTCAAC-AGACCAAGAACCTGCG 782  
DB 860 CTCACCTGCGAGGCTGTGGCTTCACTCTCTGTAAGTCAAGCAGACCAACCCCATYGG 917  
|||||

RESULT 11  
HSM805177/c 3739 bp mRNA linear HTC 20-JAN-2005  
LOCUS HSM805177  
DEFINITION Homo sapiens mRNA; cDNA DKFZp434L1127 (from clone DKFZp434L1127).  
ACCESSION AL833881  
VERSION AL833881.1 GI:21739387  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 3739)  
BLOCKER, H., BOECKER, M., BRANDT, P., MEWES, H.W., WEIL, B., AMID, C.,  
OSANGER, A., POBO, G., HAN, M. and WIEMANN, S.  
The German cDNA Consortium  
Direct Submission  
Submitted (20-JAN-2005) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuerberg, GERMANY  
COMMENT This clone (DKFZp434L1127) is available at the RZPD Deutsches  
Research Center (DKFZ); Email [s.wiemann@kfz-hidelberg.de](mailto:s.wiemann@kfz-hidelberg.de);  
Research Center (DKFZ); Email [s.wiemann@kfz-hidelberg.de](mailto:s.wiemann@kfz-hidelberg.de);  
sequenced by GSF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
This clone (DKFZp434L1127) is available at the RZPD Deutsches  
Research Center (DKFZ); Email [s.wiemann@kfz-hidelberg.de](mailto:s.wiemann@kfz-hidelberg.de);  
Please contact RZPD for ordering:  
<http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp434L1127>  
Further information about the clone and the sequencing project is  
available at <http://mips.gsf.de/projects/cdna/>.  
Location/Qualifiers  
1..3739  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp434L1127"  
/rissue\_type="testis"  
/clone\_lib="434 (synonym: hhes3). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="adult"  
/note="putative transcripct"

ORIGIN  
Query Match 21.1%; Score 421.6; DB 4; Length 3739;  
Best Local Similarity 83.8%; Pred. No. 2.4e-58;  
Matches 560; Conservative 0; Mismatches 74; Indels 34; Gaps 6;  
QY 401 TTGAGAGGTGAAGCCAGCTGGGCTT-CTGGGTCAAGTGGGAGCTTGAGAACTTTTGTGT 459  
DB 3241 TTGAGAGGTGAAGCCAACTGCTCTGGGTCCAGTGGGAGCTTGAGAACTTTTGTGT 3182  
|||||

QY	460	CTAGCTAAAGAGATTGTAAATG-----CACCATCGACCT	494
Db	3181	CTAGCTGAAGAGATTGTAAACACACCAATCAGCACTCTGTAAAAACACCAATCGACCT	3122
QY	495	CTGTGTCTAGCTAAAGAGATTGTAAATGACCAATCAGCACTCTGTAAATGGACCAATCA	554
Db	3121	CTGTGTCTAGCTAAAGAGATTGTAAACGACCAATCAGCGCTTGTAAATGGACCAATCA	3062
QY	555	GCAGGATGTGGCGGGGCTCAATTAAGGAGTAAAACTGGCCACCGGCGCAGCAGTGGC	614
Db	3061	GTAGGATGTGGCGAGGGCCAAATTAAGGAGTAAAGTGGCCACCTTGACAGTGGCGGC	3002
QY	615	AACCACTCGGGTCCCCCTTCCACACTGTGGAAAGCTTTGTTCTTTTGTCTTTCACATTA	674
Db	3001	AA-CCACTCGGGTCCCCCTTCCACACTGTGGAGCTTTGTTCTTTTGTCTTTCACAGTAA	2943
QY	675	TCTTGTCTGTCTCATTCTTTGTGTGCCACTACCTTTATGAGCTGTAACTCACTCGC	734
Db	2942	TCT---TGTGTGTCTCCTCTGTGGTCCGCACTTACCTTTATGAGCTGTAACTCACTCA	2886
QY	735	AGGCTGTGTGGCTTATTCTTCTGAAGTCAAC-AGACGAGAACCCACTGGAAGGAACAAG	793
Db	2885	AGGATCTGACGTTATTCTTGAAGTCAAGAACCAAGACCCACCGGAGGAACAAC	2822
QY	794	AACTCCGAGTGTCTGCTTTAAGAGCTGTAACTCACTCGGAGCTCTGCGACTTCA	853
Db	2825	AACTCTGAACGTCGACCTTTTAAGAGCTGTAACTCACTCGGAGGAAGTCTGAGCTTCA	2766
QY	854	TCCTGAAGTCACTGAGACCAACAACCCACAGAGGAAGAACTTGGACACACTGAT	913
Db	2765	TCCTGAAGTCACTGAGACCAATGAACCAACAGAGGAAGAACTTGGACACTTGAAC	2706
QY	914	ATCTGAAGGAACAACCTCCAGACACCACTCTTGAAGCTGTAACTCACTCCGC--AA	970
Db	2705	ATTGCAAGGAACAACCTCCGACACCACTCTTGAAGACTGTAACTCACTCCGGAAGA	2646
QY	971	GGGTCTGTGCTTCTTCTTGAAGTCAAGACCAAGAACCCACCGGAGGAACAATT	1038
Db	2645	AGGTCCGTGACTTCTTCTTGAAGTCAAGACCAAGAACCCACCGGAGGAACAATT	2586
QY	1031	CCAGACAC 1038	
Db	2585	CCGACAC 2578	
RESULT 12			
CR537266			
LOCUS	CR537266		
DEFINITION	DKR2p459N203.F1.459 (synonym: pccol1) Pongo pygmaeus cDNA clone		
ACCESSION	CR537266		
VERSION	CR537266.1		
KEYWORDS	EST.		
SOURCE	Pongo pygmaeus (orangutan)		
ORGANISM	Pongo pygmaeus		
REFERENCE	Buxarjota, Metazoa; Chordata; Vertebrate; Euteleostomi;		
AUTHORS	Mammalia; Euteria; Euarhontoglires; Primates; Catarrhini;		
	Homnidae; Pongo.		
	1 (bases 1 to 577)		
	Wambutt, R., Heubner, D., Mewes, H.W., Well, B., Amid, C., Oeangert, A.,		
	Fobo, G., Han, M. and Wiemann, S.		
	Pongo pygmaeus mRNA (Wambutt, R., Heubner, D., Mewes, H.W., et al.)		
TITLE	Unpublished (2004)		
JOURNAL	Contact: MIPS		
COMMENT			

Ingolstaedter.lnstr.1, D-85764 Neuburg, Germany. This is the 5' sequence of the clone insert. Clone from S. Wiemann Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de, sequenced by Agowa (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp559Z03) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,

Heubnerreg 6, 14059 Berlin-Charlottenburg, GERMANY; Email [clone@rpd.de](mailto:clone@rpd.de) Further information about the clone and the sequencing project is available at <http://mips.gsf.de/projects/cdna/>.

FEATURES	Location/Qualifiers
source	1. .577

```

/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="IDKP2p459N203"
/tissue_type="cortex"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="459 (synonym: pcor1)"
/note="vector: pSPORT1_S1; Site_1: S11A; Site_2: S11B"

```

## ORIGIN

Query Match	20.8%	Score 415.6	DB 7	Length 577
Best Local Similarity	87.7%	Pred. No. 3,8e-57		
Matches 465	Conservative 0	Mismatches 64	Indels 1	Gaps 1

QY	476	AAATGACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAATGACCAATCAGCACT	535
Db	48	AAATGACCAATCAGCACTCTGTGTCTAGCTAAAGGATTGTAACACACCAATCAGCACT	107

[illegible]

596 CACCCGAGCCACGAGTGGCAACCACTCGGATCCCTTCCACACTGTGGAAGCTTTGTTTC 655

Db 228 TTTCACCTTCA CATAAATCTTGCTGCTCCACCTCTTGGTCTGCACTACCTTATG 287

Ov 716 AGCTGTAACTCACTGCGAGGGTCTGTGCTTCATTTCTCTGAAGTCAAC-AGACGACGAA 774

Db 288 AGCTGTACACTCACCAGGAGGCTTGCAGCTCATTTCTGAAGTCAGACGAGACACAAA 347

Db 348 CCCACCAGAGGATAAAACACTCCAGATGACACCATGTTTAAAGGGCTATTAACACTCACTA 407

408	C	A	A	G	T	C	T	C	C	G	A	A	G	T	C	A	G	A	C	C	A	A	A	C	C	A	T	G	C	A	A	G	A	G	A	467
-----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	-----

QY 895 ACTCTGACACACCTGAATATCTGAAGAACAACTCCAGACACACCATCTTTGAGAGCT 954

Db 468 ATTCGAGACACATCTGAACATCTGAAGAACAACTCTGAGACACACCTGCTTTTAAGAGCT 527

955 GTAACTACCGCAAGGCTGTGGCTTCACTTGAAGTCAGCAAGAC 1004

## RESULT 13

CF59/0035	746 bp	mRNA	linear	EST	26-SEP-200
LOCUS					
CF59/0035					
AGENCOURT_15658367					
NICHD_Hs_ov1					
Homo sapiens					
CDNA clone					

ACCESSION	CF597035	GI:36354012
VERSION	CF597035.1	
KEYWORDS		
EST.		

SOURCE: *Homino sapiens* (humans),  
 ORGANISM: *Homo sapiens*  
 Eukaryota; Metazoa; Chordata; Vertebrata; Eureleostomi;

REFERENCE  
AUTHORS  
Hominidae; Homo.  
1 (bases 1 to 746)  
NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabds-remail.nih.gov  
Tissue Procurement: Gregory F. Erickson, Ph.D.  
cDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDCM254 row: d column: 23  
High quality sequence stop: 548.

FEATURES  
source  
1..746  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30704734"  
/tissue\_type="Ovary"  
/lab\_host="DH10B (TI phage-resistant)"  
/clone\_lib="NICHD\_Hs\_OV1"  
/note="Vector: pDNR-Lib; Site 1: SfiI (ggccatcatggcc);  
Site 2: SfiI (ggccgctcgccc); Library is oligo-dT primed  
and directionally cloned. Granulosa luteal cells aspirated  
from preovulatory follicles of normal cycling women  
undergoing ovulation induction for infertility due to male  
factor and normal donors. The cells were from follicles  
stimulated with Lupron, FSH and hCG. 5' and 3' adaptors  
were used in cloning as follows: 5' adaptor sequence:  
5'-CAGGCGCATATGATGCCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGGCCGACGAGGCGCCGACATG-dt(30)-BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 2.23  
kb (range 1.0-4.5 kb). 14/15 colonies contained inserts by  
PCR. This library was enriched for full-length clones and  
was constructed by Clontech Laboratories (Palo Alto, CA)."

## ORIGIN

Query Match 20.5%; Score 411; DB 6; Length 746;  
Best Local Similarity 85.3%; Pred. No. 1.9e-56;  
Matches 528; Conservative 0; Mismatches 60; Indels 31; Gaps 5;

```
158 TCTAGATGAGGCGCATCAAGTACAGATCTTCAATATGTAACCCCAATGAGCTCA 217
126 TATAGATGAGGCTATCAAGCTACAGATATCTTCAAAATGGAACCCCAATGAGCTCA 185
218 ACTAGCACTTCTGCTGAGGACCCCTGGAACCCGCTGGCCCTTCAATGAGCTTAAG 277
186 ACTAACAA---CTACCAAGGACCCCTGGAACCAACCCGCTGGCCCTTCAATGAGCTTAAG 242
278 AGCTCCCTCTGAGAGACACTACCACTGCAAGGCGCCCTTCTTCAACCCCTATCCAGACGA 337
243 AGTTCCCTCTGAGAGACACTACCACTGCAAGGCGCCCTTCTTGGCCCTATCCAGACGA 302
338 AGTAGCTACGACCGTATC--GCCAATCCCAACAGACGCTGGGCTGTCTTGGAGGG 396
303 AGTAGCTACGAGTGTATCAACCAATCCCAACAGACGCTGGGCTGTCTTGGAGGG 362
397 GGGATTGAGAGTGAAGCCAGCTGGGCTTCTGGGCTAGTGGGAGACTTGGAGAACTTTTG 456
363 GAAATTGAGAGTGAAGCCAGCTGGGCTTCTGGGCTGGGAGAGCTTGGAGAACTTTTC 422
457 TGTCTAGCTAAGAGATTGTAATGACCAATCAGACCTGTGTCTTGAAGTAAAGATTGT 516
423 TGTCTAGCTAAGAGATTGTAATGACCAATCAGTGTGTGTCTTGAAGTAAAGATTGT 482
517 AATTCACCAATCAGACCTCTGTAA-------TGACCA 551
483 AATTCACCAATCAGACCTCTGTAA-------TGACCA 542
```

```
QY 552 TAGACGAGTGTGGCGGCTCAATTAAGGAGTAAAGGATGGAACCCGACCGACGACT 611
DB 543 TAGATGAGTGTGGCGGCTCAATTAAGGAGTAAAGGATGGAACCCGACCGACGACT 602
QY 612 GGAACCCGACCTGGGCTGCCCTTCCACACTGT--GGAAGCTTTGTTTGGCTCTTCAACA 670
DB 603 GGAACCCGACCTGGGCTGCCCTTCCACACTGT--GGAAGCTTTGTTTGGCTCTTCAACA 662
QY 671 TAAATCTTGCTGCTGCTCAATTTCTTTGTGTCACACTACTTTATGAGCTGTAACAC-TCA 729
DB 663 TAAATCTTGCTGCTGCTCAATTTCTTTGTGTCACACTACTTTGAGTGTGACACTTCA 722
QY 730 CTGCGAAGGCTGTGTGACTT 748
DB 723 CTGCGAAGGCTTCAACTT 741
```

RESULT 14  
AQ390354/c 630 bp DNA linear GSS 06-MAR-1999  
LOCUS C17B1-E1-2545M11.TR C17B1-E1 Homo sapiens genomic clone 2545M11,  
DEFINITION genomic survey sequence.  
ACCESSION AQ390354.1 GI:4361377  
VERSION AQ390354.1  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 630)  
Zhao,S., Adam,M.D., Niernan,W., Malek,J., Shizuya,H., Simon,M. and  
Venter,J.C.  
Use of BAC End Sequences from Caltech Libraries for Sequence-Ready  
Map Building (1997)  
Unpublished (1997)  
Other GSSs: C17B1-E1-2545M11.TF  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES  
source  
1..630  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="2545M11"  
/sex="male"  
/cell\_type="sperm"  
/note="Vector: pBelobAC11; Site 1: EcoRI; Site 2: EcoRI;  
Caltech Human BAC Library D"

## ORIGIN

Query Match 20.4%; Score 408.8; DB 9; Length 630;  
Best Local Similarity 88.8%; Pred. No. 4.6e-56;  
Matches 467; Conservative 0; Mismatches 52; Indels 7; Gaps 2;

```
517 AATTCACCAATCAGACCTCTGTAATGACCAATCAGACGATGTGGCGGCTCAAA 576
624 AATTCACCAATCAGACCTCTGTAATGACCAATCAGACGATGTGGCGGCTCAAA 565
577 TAAAGGAGTAAATCTGGCCACCCGACGACGAGTGGCAACCACTCTGGGCTCCCTTCCA 636
564 TAAAGGAGTAAATCTGGCCACCCGACGACGAGTGGCAACCCGCTGGGCTCCCTTCCA 505
```

OY	637	ACTGTGGAAAGCTTTGTTCTTTTGGCTCTTCAACAATAAATCTTGGCGTCAATCTTTG	696
Db	504	CGCTTTGAAAGCTTTGTTCTTTTGGCTCTTCAATAAATCTTGGTGGCTCAATCTTTG	445
OY	697	TGTCACACTACCTTTATGAGCTGTAAACACTCACTGCGAAGGCTCTGTGGCTTCATTCCTG	756
Db	444	GGTCGCACTACCTTTATGAGCTGTAAACACTCACTGCAAGAAGTCTGTGGCTTCATTCCTG	385
OY	757	AAGTAAAC-AGACCAAGAACCCACTGTGAAGGAACAAGAATCCCGATGTGCTTCTTA	815
Db	384	AAGTGAGCAAGGCCACGAAACCCACGGGAGGAACAAACAATCCCGAAGCAACACTTTTA	325
OY	816	AGAGCTGTAAACACTCACTGCGAAGCTGTGCAAGCTTCACTCCGGAAGTCAGTGAAGACACA	875
Db	324	AGAGCTGTAAACACTCACTGCAAGGCT-----CTTCACTCCGGAAGTCAACAGACACG	271
OY	876	AAACCAACGAAGGAAGAAACTCTGGACAACACTGATATATGTGAAGGAACAAATCCACGA	935
Db	270	AAACCAACGAAGGAAGAAACTCTGGATATACATCTGAATCTTGAAGGAACAAATCCGGA	211
OY	936	CACACCAATCTTTCAAGCTGTAAACACTCAACGCAAGGCTCTGTGGCTTCATTTTGAAGT	995
Db	210	CACACCAATCTTTAAAGCTGTAAACACTCACTGCAAGGGCTCAATGCTTCATTTCTTGAAGT	151
OY	996	CAGCAAGAACAAGAACCCACGGGAAGGAACAAATTCAGACACAGT	1041
Db	150	CAGCAGAACAAGAACCCACTGGGAGAGATTAATTCGGAACATT	105

RESULT 15					
B0926428/c	B0926428	899 bp	mRNA	linear	EST 20-AUG-2002
LOCUS	AGENCOURT_8753346	Lupski_sciatic_	nerve Homo	sapiens	CDNA clone
DEFINITION	IMAGE:6205474 5',	mRNA sequence.			

ACCESSION	BO9264428	GI:22341459
VERSION	BO9264428.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens	
ORGANISM	Homo sapiens (human)	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 899)	NIH-MGC	<a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
			Contact: Robert Strausberg, Ph.D.	

Email: c9apbs-remail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MCC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LWM13628 row: g column: 11  
High quality sequence stop: 682.  
Location/Qualifiers  
1. 899

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6205474"
/sex="male"
/tissue_type="sciatic nerve"
/dev stage="adult, 70 yr"
/lab_host="DH10B"
/clone_lib="lupski_sciatic nerve"
/notes="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI, Site 2: SalI; cDNA made by oligo-dr priming.
Directionally cloned using the following adaptors:
5'-TGGACCAAGCGGTCCG-3' and

```

ORIGIN

5'-GAGCTAAGCTTAAATCGCGAGCGCGCCGCTT (15) -3'. Size selected > 1 kb for average insert length 1.87 kb. This is a primary library, non-amplified, library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

Query Match	20.1%	Score 401.6;	DB 5;	Length 899;
Best Local Similarity	87.3%	Pred. No. 6e-55;		
Matches 440; Conservative	0;	Mismatches 64;	Indels 0;	Gaps 0;

Qy	543	ATGAGCAATCAGAGGATGTGGCGGGGGTCAATAAGGAGTAAAACTGGCACCCGA	602
Db	661	ATGAGCAATCAGAGGATGTGATGGGGCAATAAGGAAATTAAGCTGGCACCCGA	602
Qy	603	GCCAGCAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGGAAGCTTGTCTTTTGT	662
Db	601	GTACGACGTGGCAACTGTGGGGTCCCTTCCACGCTGGGAAGCTTGTCTTTGTCT	542
Qy	663	CTTACACATAATCTTGTCTGTGCTCAATCTTTGTGTCCACACTACTTTATGAGCTGA	722
Db	541	CTTACATAATGATTTTGTCTGTGCTCACTTTTGGGTCCGCACTGCTTTATGAGCTGA	482
Qy	723	ACACTCACTCGAGGGTCTGTGGCTTCAATTCCTGAATCAACGACCAAGACCCACTGG	782
Db	481	ACACTCACAGAGGGTCTGGGCTTCAATTCCTCAATGAGCAAGACCAACCCACTGG	422

Qy	783	AAGGACAAAAGATCTCCGAGTGTGCTGCTTTAAGACTGTAACTACTGCGAGCTC	843
Db	421	GAGGACAAAACAACTCCAGACGGGCCACTTTTAAAGACTGTAACTACTGCGAAGCTC	362
Qy	843	TGCAGCTTCACTCCTGAAGTCAGTGGAGACCAACAACCCACAGAAAGGAAAGAACTCGGA	902
Db	361	CGCTGCTTCACTCCCGAAGTTAAGGAGACCAACCAATGAAAGGAAAGAACTCCGGA	302
Qy	903	CACACTGAATATCTGAAGAAACAAATCTCCAGACACACACTCTTTAGACTGTAACT	962
Db	301	CACATCTGAACATCTGAAACAAACAAATCTCCAGACACACACTCTTTAAGACTTTAACATT	242
Qy	963	CACCGAAGGCTCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAG	1022
Db	241	CACGCGCAAGGCTCTGCGGCTTCAATCTTGAAGTCAGCAAGAACCAATGGAAG	182
Qy	1023	AAACAATTCCAGACACAGTAGGAA	1046
Db	181	AATTAATTCCAGACACATCGGAA	158

RESULT 16	
LOCUS	BO678014
DEFINITION	BO678014 900 bp mRNA linear EST 15-JUN-2002
ACCESSION	AGNCNCOURT_8033102 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6091116
VERSION	5, mRNA sequence.
KEYWORDS	BO678014 BO678014.1 GI:21790693
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 900)	NIH-MGC	<a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a>	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.			
	Email: <a href="mailto:cgabs-r@mail.nih.gov">cgabs-r@mail.nih.gov</a>			
	Tissue Procurement: DCTD/DTF			
	CDNA Library Preparation: Rubin Laboratory			
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			

Found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM233 row: j column: 13  
High quality sequence stop: 652.  
Location/Qualifiers

## FEATURES

source

1..900

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6091116"  
/tissue\_type="melanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOT7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 19.9%; Score 397.2; DB 5; Length 900;  
Best Local Similarity 84.2%; Pred. No. 3.1e-54;  
Matches 507; Conservative 0; Mismatches 63; Indels 32; Gaps 4;

```

OY 154 TTCTCTAGAGTGGGCGCATCAAGCTACAGATGATTTCAATGTAACCCCAATGAG 213
DB 252 TTATTTGATGAGGAGGCTATCAAGCTACAGATGATTTCAATGTAACCCCAATGAG 311
OY 214 CTCACCTAACACTTCTGCTGAGGACCCCTGACCCGCTGGCCCTTTCATGAGGCT 273
DB 312 CTCACCTAACAAA---CTACCAAGGACCCCTGACCAACCCGCTGGCCCTTTCATGAGGCT 368
OY 274 AAAGAGTCCCTCTGAGGACACTACACTGAGGCGCCCTTCTTCAACCCCTATCCAGC 333
DB 369 AAAGAGTCCCTCTGAGGACACTACACTGAGGCGCTTCTTTCATGAGGCT 428
OY 334 AGAAGAGCTACAGGCGGTGATC-GCCAAATCCCAACAGAGCGGGGTCCTGTTTGG 392
DB 429 AGAAGAGCTACAGGCGGTGATCAGCCCAATCCCAACAGAGTTGGGTCCTGTTTGAAG 488
OY 393 AGGCGGATTTGAGAGGTGAAGCAGCTGGCTTCTGGGTGAGGTGGGACTTGGAGAACT 452
DB 489 TGGGAGATTTGAGAGGTGAAGCAGCTGGCTTCTGGGTGAGGTGGGACTTGGAGAACT 548
OY 453 TTTGTGTCTGCTAAAGGATTTGAATGCAACATCAGCACTGTGTCTTAAAGCA 512
DB 549 CTCTGTCTGCTGAGGATTTGAATGCAACATCAGTGTGTCTTAAAGCACTGAGGT 608
OY 513 TTGTTAATGC-----ACCAATCAGCACTGTGTTAATG 546
DB 609 TTGTTAATGCAACATCAGCACTGTGTTAATGCACTGTGTTAATG 668
OY 547 ACCAATCAGCAGGATGTGGCGGGGTCAAATTAAGGAGTAATAATGCGCACCCGAGCA 606
DB 669 ACCAATCAGGATGTGGCGGGGTCAAATTAAGGAGTAATAATGCGCACCCGAGCA 728
OY 607 GCAGTGGCAACCACTGGGTCCTTTCACACTGTGGAAGCTTTGTTTCTTCTTTC 666
DB 729 GCAGTGGCAACCACTGGGTCCTTTCACACTGTGGAAGCTTTGTTTCTTTCACCTTTC 788
OY 667 ACAATTAATCTGTGTCTGCTCACTTTT--GTGTCACTACTACTTATGAGCTGTAAC 724
DB 789 ACAATTAATCTGTGTCTGCTCACTTTG3GGGCCCCCACCACCTTTATGAGCTGGAC 848
OY 725 AC 726
DB 849 AC 850
```

RESULT 17

BU602065 816 bp mRNA linear EST 20-SEP-2002  
LOCUS  
AGENCOURT 10015763 NIH\_MGC\_142 Homo sapiens cDNA clone  
IMAGE:6496097 5', mRNA sequence.

ACCESSION  
BU602065  
VERSION  
BU602065.1 GI:23253824  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
NIH-MGC http://mhc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: NCI  
CDNA Library Preparation: Michael Brownstein Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2674 row: p column: 18  
High quality sequence stop: 635.

## FEATURES

source

1..816  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/clone="IMAGE:6496097"  
/tissue\_type="mixed (pool of 40 RNAs)"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_lib="NIH\_MGC\_142"  
/note="Vector: pDNR-LIB; Site 1: SfiI (ggccatcggcc);  
Site 2: SfiI (ggccgcctcgcc); Double-stranded cDNA was  
prepared from a pool of 40 cell line polyA+ RNAs (bladder  
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -  
4%, connective tissue - 1.4%, eye - 1% intestine - 2.6%,  
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,  
ovary - 1.3%, pharynx - 2.5%, prostate - 4.3%, salivary  
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were  
used in cloning as follows:  
5'-AAGCAGTGTATCAAGCAGAGTGGCATTTACGGCCGG-3' and  
5'-ATTCTAGAGCCGAGCGGCCACATG-dT(30)NN-3'. Full-length  
enriched library was constructed using the Clontech  
Creator SMART kit and size-selected to contain the >0.5 kb  
size fraction (other fractions present in NIH\_MGC\_141).  
Library created in the laboratory of M. Brownstein (NIH,  
NIH). Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 19.6%; Score 392.8; DB 5; Length 816;  
Best Local Similarity 84.2%; Pred. No. 1.6e-53;  
Matches 500; Conservative 0; Mismatches 64; Indels 30; Gaps 4;

```

OY 161 AGATGAGGCGCATCAAGTACAGATGATTTCAAAATGTAACCCCAATGAGCTCAAT 220
DB 217 AGATGAGGCGCATCAAGTACAGATGATTTCAAAATGTAACCCCAATGAGCTCAAT 276
OY 221 AACAACTTGTGCTGAGGACCCCTGACCGAGCCGCTGGCTTTCAATGAGCTTAAAGAGC 280
DB 277 AACAA---CTACCAAGGACCCCTGAGCAACCCGCTGGCCCTTTCATGAGCTTAAAGAGT 333
OY 281 TCCCTCTGAGAGCACTACCACTGAGGCGCCCTTTCACCCCTATCAGCAGAACT 340
DB 334 TCCCTCTGAGAGCACTACCACTGAGGCGCTTTCCTTTCGCCCCCTATCAGCAGAACT 393
OY 341 AGCTACAGCGGTTCATC-GCCAAATCCCAACAGAGTGGGTCCTGTTTGAAGGAGGAG 399
DB 394 AGCTACAGGTTCATCAGCAGCAATCCCAATCCCAACAGAGTGGGTCCTGTTTGAAGGAGGAG 453
```



QY 400 ATTGAGGAGTGAAGCCAGCTGGGCTTCTGGGTGACAGTGGGGAAGCTTGGAAGACTTTTGT 459  
DB 454 ATTGAGGAGTGAAGCCAGCTGGGCTTCTGGGTGAGGAGTGGGAAGACTTTTGT 513  
QY 460 CTAGCTAAAGGATTTGTAATGACCAATCAGCACTGTGTCTAGTAAAGATTTGTA 519  
DB 514 CTAGCTAAGGATTTGTAATGACCAATCAGCTGTGTCTAGTAAAGATTTGTA 573  
QY 520 TGC-----ACCAATCAGCACTGTGTAAATGACCAATCA 554  
DB 574 TGACCAATCAGCACTGTGTAAATGACCAATCAGCACTGTGTAAATGACCAATCA 633  
QY 555 GCAGAGATGTGGGGGGTCAATTAAGGAGTAAATATGCGCAACCCAGCAGAGTGC 614  
DB 634 GTAGGATGCGGCGCAATTAAGGAGTAAATGAGTGCACCTGAGTCAAGTGC 693  
QY 615 AACCACTCGGGGTCCTTCCACACTGTGAGAGCTTTGTTCTTTGCTTTCAATTA 674  
DB 694 AACCACTCGGGTCCCTTTCACTGTGTGAGAGCTTTGTTCTTTCACTTTCAATTA 753  
QY 675 TCTTGTGCTGCTCATTCTTTGTGTGTCACACTA-CCTTTATGAGCTGTAAACT 727  
DB 754 CTGCTGCTGCTCATTCTTTGTGTGTCACACTA-CCTTTATGAGCTGTAAACT 807

RESULT 18  
CD685118/c 635 bp mRNA linear EST 25-JUN-2003

LOCUS CD685118  
DEFINITION EST1638 human nasopharynx Homo sapiens cDNA, mRNA sequence.  
ACCESSION CD685118  
VERSION CD685118.1 GI:32200775  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
AUTHORS Liu,X.-Q., Zhou,Y., Zhang,J.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and  
Zeng,Y.-X.  
TITLE Transcriptional Gene Expression Profile of Human Nasopharynx  
JOURNAL Unpublished (2003)  
COMMENT Contact: Yixin Zeng  
Cancer Center  
Sun Yat-sen University  
651 Dongfeng Road East, Guangzhou 510660, China  
Tel: 86-1380-9770-743  
Fax: 86-20-8775-4506  
Email: yxzeng@zsums.edu.cn.

FEATURES  
source Location/Qualifiers  
1..635  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/issue\_type="normal nasopharynx"  
/clone\_lib="human nasopharynx"  
/note="ESTs generated from a normal nasopharynx cDNA  
library from southern Chinese"

ORIGIN  
Query Match 19.5%; Score 390.6; DB 6; Length 635;  
Best Local Similarity 86.4%; Pred. No. 3.9e-53;  
Matches 443; Conservative 0; Mismatches 69; Indels 1; Gaps 1;  
QY 530 AGCACTCTGTAATGACCAATCAGCAGATGTGGCGGGTCAATTAAGGAGTAA 589  
DB 635 AGCTCTCTGTAATGATCAATCAGCAGATGTGGGTGGGCGCAATTAAGGAGTAA 576  
QY 590 ACTGGCCACCGAGCCAGCAGTGGCAACCACTCGGGTCCCTTCCACACTGTGAAGT 649  
DB 575 GCAGTCCACCTGAGCCAGCAGTGGCAACCTCGGTCCAGTCCCTTCCAGCTGTGAAGT 516

QY 650 TTGTTCTTTGCTCTTCAATTAATCTTGCTGCTCTATCTTTGTGTCACACTACC 709  
DB 515 TTGTTCTTTGCTCTTCAATTAATCTTGCTGCTCTATCTTTGTGTCACACTACC 456  
QY 710 TTTATGAGCTGTAACTCACTGCGAGGCTGTGTGCTTCAATCTGGAAGTAAAC-AGAC 768  
DB 455 TTGTGAGCTGTAACTCACTGCGAGGCTGTGTGCTTCAATCTGGAAGTAAAC-AGAC 396  
QY 769 CAGGACCCACTGGAAGGAAACAAAGATCCCGATGTGCTGCTTTAAGAGCTGTAAAC 828  
DB 395 CATGAACCCACAGGAGGAAATGAACATCTGACATCCCGCTTTAAGAGCTGTAACT 336  
QY 829 TCACCTGAGAGCTCTGAGCTTCACTCTGGAAGTCAAGTGAAGCAACCAACCAAG 888  
DB 335 TCACCTGAGAGCTCTGAGCTTCACTCTGGAAGTCAAGTGAAGCAACCAACCAAG 276  
QY 889 GAAGAACTCTGAGCACTGAGTGAATATCTGAAGAAACAACTTCAGACACACCATCTTT 948  
DB 275 GAAGAACTCTGAGCACTGAGTGAATATCTGAAGAAACAACTTCAGACACACCATCTTT 216  
QY 949 AGAGCTGTAACTCACTGCGAGGCTGTGTGCTTCAATCTGGAAGTGAAGCAACCAAG 1008  
DB 215 AGAGCTGTAACTCACTGCGAGGCTGTGTGCTTCAATCTGGAAGTGAAGCAACCAAG 156  
QY 1009 AACCCACCGGAGGAAACAAATTCAGACACAGT 1041  
DB 155 AACCCACCGGAGGAAACAAATTCAGACACAGT 123

RESULT 19  
AL704833/c 617 bp mRNA linear EST 04-SEP-2003  
LOCUS AL704833  
DEFINITION DKFZp686B2033\_r1.686 (synonym: h1cc3) Homo sapiens cDNA clone  
ACCESSION AL704833  
VERSION AL704833.1 GI:19688188  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and  
Wiemann,S.  
TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and  
Wiemann,S.)  
JOURNAL Unpublished (2001)  
COMMENT Contact: MIPS

FEATURES  
source MIPS  
1..617  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp686B2033"  
/dev stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="686 (synonym: h1cc3)"  
/note="Vector: pTriplex2; Site\_1: SfiI; Site\_2: SfiI; cDNA-collection"

ORIGIN  
Query Match 19.5%; Score 390.6; DB 6; Length 635;  
Best Local Similarity 86.4%; Pred. No. 3.9e-53;  
Matches 443; Conservative 0; Mismatches 69; Indels 1; Gaps 1;  
QY 530 AGCACTCTGTAATGACCAATCAGCAGATGTGGCGGGTCAATTAAGGAGTAA 589  
DB 635 AGCTCTCTGTAATGATCAATCAGCAGATGTGGGTGGGCGCAATTAAGGAGTAA 576  
QY 590 ACTGGCCACCGAGCCAGCAGTGGCAACCACTCGGGTCCCTTCCACACTGTGAAGT 649  
DB 575 GCAGTCCACCTGAGCCAGCAGTGGCAACCTCGGTCCAGTCCCTTCCAGCTGTGAAGT 516

ORIGIN  
Query Match 19.5%; Score 390.6; DB 6; Length 635;  
Best Local Similarity 86.4%; Pred. No. 3.9e-53;  
Matches 443; Conservative 0; Mismatches 69; Indels 1; Gaps 1;  
QY 530 AGCACTCTGTAATGACCAATCAGCAGATGTGGCGGGTCAATTAAGGAGTAA 589  
DB 635 AGCTCTCTGTAATGATCAATCAGCAGATGTGGGTGGGCGCAATTAAGGAGTAA 576  
QY 590 ACTGGCCACCGAGCCAGCAGTGGCAACCACTCGGGTCCCTTCCACACTGTGAAGT 649  
DB 575 GCAGTCCACCTGAGCCAGCAGTGGCAACCTCGGTCCAGTCCCTTCCAGCTGTGAAGT 516

ORIGIN  
Query Match 19.5%; Score 390.6; DB 6; Length 635;  
Best Local Similarity 86.4%; Pred. No. 3.9e-53;  
Matches 443; Conservative 0; Mismatches 69; Indels 1; Gaps 1;  
QY 530 AGCACTCTGTAATGACCAATCAGCAGATGTGGCGGGTCAATTAAGGAGTAA 589  
DB 635 AGCTCTCTGTAATGATCAATCAGCAGATGTGGGTGGGCGCAATTAAGGAGTAA 576  
QY 590 ACTGGCCACCGAGCCAGCAGTGGCAACCACTCGGGTCCCTTCCACACTGTGAAGT 649  
DB 575 GCAGTCCACCTGAGCCAGCAGTGGCAACCTCGGTCCAGTCCCTTCCAGCTGTGAAGT 516





Db	474	ATTAATCTTGCTGCTCTCACTCTTCGGGTCACATGCTCTTTAAGAGTGAACCTCA	415
QY	730	CTGCAGAGGCTCTGCGCTTCACTTCCGAAAGTCAC-AGACCAAGAACCCACTGGAAGAA	788
Db	414	CCCCGAAGTCTGAAGCTTCACTCCTGAAAGCAGAGACCGCGAACCACCTGGGAGGAA	355
QY	789	CAGAAGACTCCCGCATGTGCTGCTTTAAGAAGTTAAACACTCACTGCGAAGCTCTGCAGC	848
Db	354	TGAACMACTCCAGAGTGCCTGCGC-TTAAAGCTGTAAACCTCACTGCGCAAGGCTGCAAGC	296
QY	849	TTCACTCCTGAAGTAGTGAGACCAAAACCACGAAAGGAAGAACTGTGACACACC	908
Db	295	TTCACTCTGCTG-AGCCAGCGAGACACGAAACCACAGAAAGTAAGAACTCTGAACACTC	237
QY	909	TGAATATCTGAAGGAACAAACTCCAGACACACCACTCTTTCAGAGCTGTAAACCTCACCGC	968
Db	236	CGAACATCAGAAAGAACAAACTCTGGAATGCTGCTTTAAGAACTGTAACTCTTAACCGC	177
QY	969	AAGGCTCTGCTTCACTTTTGAAGTCAGCAAGACCAAGAACCCACC	1016
Db	176	AAGGCTCGCGCTTCACTTTTGAAGTCAGTAGACCAAGAACCCACC	129

RESULT 23				
CD242806				
LOCUS	929 bp	mRNA	linear	EST 22-MAY-2003
DEFINITION	CD242806			
	AGENCOURT_14126047	NIH_MGC_179	Homo sapiens	cDNA clone

ACCESSION	CD242806	GI:31003270
VERSION	CD242806.1	
KEYWORDS	EST.	

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
1 (bases 1 to 929)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC).  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strusberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MSC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Plate: NDAM448 row: d column: 10  
High quality sequence start: 4  
High quality sequence stop: 499.

FEATURES	Location/Qualifiers
source	1. .929

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30383793"
/tissue_type="Pituitary"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_179"
/note="Organ: brain; Vector: pCMV-SPORT6.1; Site_1: EcoRV
(destroyed); Site_2: NotI; library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.1 kb. library was
constructed by (Invitrogen). Note: this is a NIH_MGC
library."

```

ORIGIN	
Query Match	18.5%; Score 370; DB 6; Length 929;
Best Local Similarity	84.8%; Pred. No. 7.4e-50;
Matches 451; Conservative	0; Mismatches 75; Indels 6; Gaps 3

QY	516	TAATATGACCAATAGCACTCTGTAAATATGGACCAATAGCAGGAATGTGGCGGGCTTCA	575
Db	2	TAATATGACCAATAGCTGCTCTGTAAATATGACCAAGTCAAGCAAGATGTGGCGAGGCCAA	61
QY	576	ATAAGGAGTAAATAATGAGCCACCCGAGCCAGAGTGGCAACCCACTGGGATCCCTTCC	635
Db	62	ATAAGGAGTAAATAATGAGCCACCCGAGCCAGAGTGGCAACCCACTGGGATCCCTTCC	117
QY	636	AACATGTGGAAGCTTGTCTCTTTTGTCTGTCAATTAATCTTGGCTGCTCATTTCTT	695
Db	118	AACATGTGGAAGCTTGTCTCTTTTGTCTGTCAATTAATCTTGGCTGCTCATTTCTT	177
QY	696	GTGTCCACACTACCTTTATATGAGCTGTAAACATCACTGCGAGGGTCTGTGACTTACTT	755
Db	178	GGGTGTGCGCTGCTTGTGTGAGCTGTGCACTCACTGCGAGGGTCTGTGACTTACTT	237
QY	756	GAAATGACAC - AGACCGAGAACCCATGTGAAGGAACAAAGAACTCCGATGTGCTGCTT	814
Db	238	GAAATGACAGAACCCAAACCCACTGTGGAGAGAAACAAACATTTAGAGTGTCCACTTT	297
QY	815	AAGAGCTTAAACATCACTGAGGAACTGTGAGCTTCACTCCTGAAGTC - AGTGAAGCA	873
Db	298	AAGAGCTTAAACATGTGCTGCGAAGGTCTGTGGCTTCACTCCTGAAGTCAACCGAGCCA	357
QY	874	CAAAACCCACCGAAGAGAAACCTGTGACACACCTGAATATCTGAAGGAACAAACTCCA	933
Db	358	TGAACCCACCGAAGAGAAACCTGTGACACACCTGAATATCTGAAGGAACAAACTCCG	417
QY	934	GACACACCATCTTTAGAGCTGTAAACACTCACCGGAAGGGTCTGTGGCTTCAATTTGAA	993
Db	418	GACACACCATCTTTAGAGCTGTAAACACTCACCGAGGGTCCATGGCTTCGTTCTTGAA	477
QY	994	GTACAGCAAGATCCAAAGACCCGAGAGAAACAAATTTCCAGACACAGTGA	1045
Db	478	GTACAGAGACCCAAAGACCCGAGAGAAACAAATTTCCGACACACTAGTA	529

RESULT 24				
BZ600660/c				
LOCUS	BZ600660	821 bp	DNA	linear
DEFINITION	WHAB7ATF Human MCF7 breast cancer cell line library (MCF7_1) Homo sapiens genomic clone MCF7_1-TM4, genomic survey sequence.			

ACCESSION	BZ600660	GI:31509122
VERSION	BZ600660.1	
KEYWORDS	GSS.	

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
1 (bases 1 to 821)

AUTHORS	Volk, S., Zhao, S., Chin, K., Brebner, J. H., Herndon, D. R., Tao, Q., Kowbel, D., Huang, G., Lapuk, A., Xu, W.-L., Magrane, G., de Jong, P., Gray, J. W. and Collins, C.
TITLE	End-sequence profiling: Sequence-based analysis of aberrant genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
PMID	12788976
COMMENT	Contact: Volk SV

Colin Collins' lab  
UCSF Comprehensive Cancer Center  
UCSF Box 0808, San Francisco, CA 94143-0808, USA  
Tel: 415 502 7066  
Fax: 415 502 5665  
Email: [svollk@cc.ucsf.edu](mailto:svollk@cc.ucsf.edu)  
This clone is available from Amplicon Express  
<http://www.genomex.com>  
Class: BAC ends.

```
FEATURES
  source
    location/Qualifiers
      1..821
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
```



QY	734	GAGGCTCGTGGCTTCATCTCTGGAATCAAGACGACGAAACCCACTGGAGAACAAG	734
Db	794	AAAGTCTTCGACTTACTCTCTTAAGGTACGAGACCAAGACCATGGAGGACAAATCG	853
QY	794	A 794	
Db	854	A 854	
RESULT 26			
BQ680536			
LOCUS			
DEFINITION	BQ680536	937 bp	mRNA
ACCESSION	AGENCOCURT 8507906	NIH_MGC_112	Homo sapiens cDNA clone IMAGE:6294411
VERSION	5', mRNA sequence.		
KEYWORDS	BQ680536	1	GI:21793215
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euteria; Euarchontoglires; Primates; Catarrhini;		
JOURNAL	Homnidae; Homo.		
COMMENT	1 (bases 1 to 937)		
	NIH-MGC http://mhc.nci.nih.gov/.		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-r@mail.nih.gov		
	Tissue Procurement: DCTD/DTF		
	cDNA Library Preparation: Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: L16C2499 row: a column: 06		
	High quality sequence stop: 103.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..937		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6294413"		
	/tissue_type="melanotic melanoma, cell line"		
	/lab_host="PH10B (phage-resistant)"		
	/clone_1ib="NIH MGC 112"		
	/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:		
	ECORI; cDNA made by oligo-dT priming. Directionally clone		
	into EcorI/XhoI sites using the following 5' adaptor:		
	GGCAGAG(G). Library constructed by Ling Hong in the		
	laboratory of Gerald M. Rubin (University of California,		
	Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and		
	Superscript II RT (Life Technologies). Note: this is a		
	NIH_MGC Library."		
ORIGIN			
Query Match	18.5%;	Score 369.2;	DB 5; Length 937;
Best Local Similarity	82.9%;	Pred. No. 9.9e-50;	
Matches 528; Conservative	0; Mismatches 74;	Indels 35; Gaps 8	
QY	158	TTTAGAGTGGAGGACCTCAAGCTACAGATGATTTTACAATGTAACCCCAAGACTCA	217
Db	274	TTTGGAGTGAAGGCTATCAAGCTACAGATGATTTTCAATGGAACCCCAAGACTCA	333
QY	218	ACTAACCACTTCGTGCTGAGGAGCCCTCGAACCCGCTGGCCCTTTCATGTGACTAAAG	277
Db	334	ACTAACAA--CTACCAAGGACCCCTTGACCAACCCGCTGGCCCTTTCATGTGACTAAAG	390
QY	278	AGCTCCCTCTGGAGGACACTACCACTGACGGGCCCTTCTTTCACCCCTATCCAGAGGA	337
Db	391	AGTTCCCTCTGGAGGACACTACCACTGAGGGGTCTTTCTTTGGCCCTATCCAGAGGA	450
QY	338	AGTAGCTACAGCGGTATC-GCCAAATCCCAACAGACGCTGGGGGTGCTCTTTTGGAGGG	396

Db 451 AGTAGCTAAGTGTGCATCACCCAACTTCCAAAGCAGTTGGGGTGTCTTTAAATGGG 510

Qy 397 GGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAAGTGGGGACTTGGAACTTTTG 456

Db 511 GAAATTGAAGGTGAAGCCAGCTGGGCTTCTGGGTGGGGACTTGGAACTCTTTC 570

Qy 457 TGTCTAGCTAAAGATTGTAATGACCAATCAGACTCTGTGTCTAGCTAAAGATTGT 516

Db 571 TGGCTATGCTAGAGATTGGAAACAACAACCAATCACTGTATGTGTCTAGCTAGAGGGTGTG 630

Qy 517 AAATGC-----ACCAATCAGACTCTGTAAATGGACAA 551

Db 631 AAATGCAACCAATCAGACTCTGTAAAAACGACCAATCAGACTCTGTAAATGGACAA 690

Qy 552 TCAGCAGATGTGGCGGGGTCAATTAAGGATTAAGCTGGCAACCCAGACAGT 611

Db 691 TCGTATGATGTGGGCAAGGGCAATTAAGGA-TAANAAGCTGGCACTATAGTCAAG 749

Qy 612 GGCACCACTCGGGTCCCTTCACACTGT-GGAAGCTTGTCTTTGCTCTTACA 670

Db 750 GGCACCACTCGGGTCCCTTCATGTGTGGGAACCTTTGGCTTTCACCTTTAC-A 808

Qy 671 TAAATCTTGTCTGTCTCATTTCTTTGTGTCCACACTA-CCTTTATAGCTGTAACTCA 729

Db 809 TAAATCTTGTCTGTCTCATCTTTTGGGGCCACACACCCTTTATAGCTGCAACTCA 868

Qy 730 CTG--CGAGGCTGTGTGCTTCATTCGTAAGTCAAC 764

Db 869 CTGGCGAAGGGCTCTCAGCTTCACTCTCGAAGGGCAGC 905

```

RESULT 27
LOCUS BU168299          889 bp      mRNA           linear   EST 04-SEP-2000
DEFINITION BUENECOURT_7941328 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011576
            5', mRNA sequence.
ACCESSION  BU168299
VERSION    BU168299.1 GI:22682283
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
             1 (bases 1 to 889)
REFERENCE  NIH-MGC http://mgs.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
AUTHORS   Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue Procurement: DCTD/DTP/gasdar CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLt) DNA Sequencing By: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLT at: http://image.llnl.gov Plate: LLAM3201 row= h column= 09 High quality sequence stop. 615.
FEATURES
         Location/Qualifiers
                1..889
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:6011576"
                 /cistue_type="large cell carcinoma"
                 /lab_host="DHIOB [phage-resistant]"
                 /clote_lib="NIH MGC 68"
                 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dR Technologies." Average insert size 1.8 kb. Library constructed by Life

```





REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 789)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genetics / NIH  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: James Martin, University of Iowa  
CDNA Library Preparation: M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: NDAMI071 row: j column: 02  
High quality sequence stop: 695.

FEATURES  
source

Location/Qualifiers  
1..789  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30707257"  
/lab\_host="DH10B Tona"  
/clone\_lib="NIH\_MGC\_220"  
/note="Organ: mixed; Vector: pYX-Asc; Site: 1: EcoRI;  
Site 2: NotI; Library is oligo-dT primed and directionally  
cloned. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to RNA size fraction, ligated with  
EcoR I adaptor, digested with Not I and then cloned  
directionally into pYX-Asc vector. Average insert size  
0.5-1kb. Adaptors 5' (AATTCGACGAGG) 3' and 5' d  
(CCTGTCGCG) 3'. 3' linker sequence - GGCGCGCTGAGAGCC T18.  
Sequencing primers 3' end: T3 promoter primer 5' d  
(ATTACCTCTCATTAAGGA) 3'. 5' End: T7 promoter primer 5' d  
(TAATCGACTCACTAATGAG) 3'. Library was constructed in the  
laboratory of M. Bento Soares. Average insert size 3-4kb  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 18.2%; Score 365; DB 6; Length 789;  
Best Local Similarity 85.8%; Pred. No. 4.9e-49;  
Matches 452; Conservative 0; Mismatches 70; Indels 5; Gaps 4;

QY 517 AATGACCAATAGAGACTGTAAATGGAACCAATCAGAGATGGGCGGGTCAAA 576  
DB 42 AAGTGACCAATAGAGGCTGTAAATGGAACCAATCAGATGTGGCGGGCCAAA 101  
QY 577 TAAAGAGTAAAAAAGTGGCCAGCCGAGCGAGTGGCAACCACTCGGGTCCCTTCA 636  
DB 102 TAAAGGAATAAAGCTGACACATGAGCGAGCAACGACCCGCTGGGCTGCCCTTCA 161  
QY 637 CACTGTGGAAGCTTTGTTCTTTTGTCTTCAATTAATTTGCTGTCTCAATTTCTTG 696  
DB 162 CCTCGTGGGGCTTTGTTGTTTCATTCATCAATTAATCTGTTGCTCACTGTTTG 221  
QY 697 TGTCCACACTACCTTATAGCTGTAACTCACTGCGA--GGGTGTGGCTTCATTC 754  
DB 222 GGTCCACACTACCTTATAGCTGTAACTCCAGGAGTGGGTCTGTGGCTTCATTC 281  
QY 755 TGAAGTCAA-CAGACACGAACCACTGGAAGAAACAAGAACTCCGATGTGCTGCTT 813  
DB 282 TGAAGTCAAGAGACCACTGAACCTCTGGAGAGGAAACAACACTCTGGATGCCACTT 341  
QY 814 TAAAGCTGTAACTCTGAGAGCTGTGCAAGCTTCACTCTGAAATCAGTGAACCA 873  
DB 342 TTAGAGCTGAACCTCTCACTG-GAAGGTCCACAGCTTCATTTGAAGTACGAGACCA 400

QY 874 CAATCCACAGAAAGAAACTCTGGA-CACACCTGAATATCTGAAGAACAACTCC 932  
DB 401 TGAACCCACCGAAGAAAGAACTCCGATACACTGTGAACATCTGAAGAACAACTCT 460  
QY 933 AGACACACCATTTTCAAGCTGTAACTCACTCCGCAAGGCTGTGCTTCACTTTTGA 992  
DB 461 GGACACACCATTTTAAAGGCTGTAACTCACTCTGTAAGGTCTCGGCTTCATTTTGA 520  
QY 993 AGTCAGCAAGACCAAGAACCCACCGAAGAAACAATTCAGACACA 1039  
DB 521 AGTCAGCCAGACCAAGAACCACTTGAAGAAATTAATTCGAGACACA 567

RESULT 30  
AO839854  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AO839854 14771 bp DNA linear GSS 31-MAR-2000  
260L13-C54 C1TB Homo sapiens genomic clone 260L13, genomic survey  
sequence.  
AO839854  
AO839854.1 GI:6652486  
GSS.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
COMMENT

1 (bases 1 to 14771)  
Carten,J.D., Makalowska,I., Robbing,C.M., Scott,N., Sood,R.,  
Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,  
Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,  
Williams,H., McDonald,L., Baxevanis,A.D., Klingler,K.W. and  
Landes,G.M.  
A 6-Mb high-resolution physical and transcription map encompassing  
the hereditary prostate cancer 1 (HPC1) region  
Genomics 64 (1), 1-14 (2000)  
10708513  
Contact: Carten JD  
Cancer Genetics Branch  
National Human Genome Research Institute/National Institutes of  
Health  
Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD  
Tel: 301 435 5626  
Fax: 301 435 5465  
Email: [jdc@nhgri.nih.gov](mailto:jdc@nhgri.nih.gov)  
Class: Shotgun.

## FEATURES

source  
Location/Qualifiers  
1..14771  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="260L13"  
/clone\_lib="C1TB"

## ORIGIN

Query Match 18.1%; Score 363; DB 9; Length 14771;  
Best Local Similarity 81.1%; Pred. No. 4.4e-49;  
Matches 481; Conservative 0; Mismatches 85; Indels 27; Gaps 4;

QY 453 TTTGTGTAGCTAAAGATTGTAAATGCAACAATCAGACTCTGTGTAGCTAAAGGA 512  
DB 1665 TCTGTGTAGCTAAAGATTGTAAATGCAACAATCAGACTCTGTGTAGCTAAAGGA 1712  
QY 513 TTGTAAATGACCAATCAGACTCTGTAAATGCAACAATCAGAGATGGGCGGGGT 572  
DB 1713 ---AAATGACCAATCAGACTCTGTAAATGCAACAATCAGATGGGCGGGGT 1768  
QY 573 CAATTAAGGAGTAAAACTGGCCACCCGAGCAGAGTGGCAACCACTCCGGTCCCT 632  
DB 1769 CAATTAAGGAGTAAAAAGTGGCCACCCGAGCAGAGTGGCAACCACTCTTGTGTCAC 1828  
QY 633 TCCACACTGTGAAGCTTTGTTCTTTGCTCTTCACAATTAATCTTGTGCTGCTCATTC 692  
DB 1713 TTTGTGTAGCTAAAGATTGTAAATGCAACAATCAGACTCTGTGTAGCTAAAGGA 512

Db 1829 TCTATGCTGTGGAAGCTTTGTTGTTGCTCTCACAATAATCTTCTGCTGCACTC 1888  
QY 693 TTTGTGTCAC---ACTACTTTATAGAGTGTAACTCACTGAGGGGTCTGTGCTTC 749  
Db 1889 TTTGGGTCGCACTCACTCTTATGAGCTGTAACTCACTCACTGAGGGGTCTGAGCTTC 1948  
QY 750 ATTCTGTAAGTCA-ACAGACCAAGAAACCACTGGAAGAAACAAAGAACTCCGATGCT 808  
Db 1949 ATTCTGTAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2008  
QY 809 GCTTTAAGAGCTGTAACTCACTGAGAGCTCTGCACTCTCTGTAAGTCACTGTA 868  
Db 2009 ACCTTTAAGAGCTGTAACTCACTGAGAGCTCTGCACTCTCTGTAAGTCACTGTA 2068  
QY 869 GACCAACAAACCCCAAG 928  
Db 2069 GACCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 2121  
QY 929 CTCAGACACACCATCTTTGAGAGCTGTAACTCACTCACTCACTCACTCACTCACT 988  
Db 2122 TTTCCGACACACCATCTTTAAGAGCTGTAACTCACTCACTGTAAGTCTGTGCTTC 2181  
QY 989 TTGAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 1041  
Db 2182 TTGAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 2234

RESULT 31  
AQ742559/c 764 bp DNA linear GSS 16-JUL-1999  
LOCUS HS.5384.B1.B11.SP6.RPCT-11 Human Male BAC Library Homo sapiens  
DEFINITION genomic clone Plate=960 Col=21 Row=D, genomic survey sequence.  
ACCESSION AQ742559  
VERSION AQ742559.1 GI:5520081  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 764)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCT-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 960 row: D column: 21  
Seg primer: SP6  
Clase: BAC ends  
High quality sequence stop: 764.  
Location/Qualifiers  
1..764  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=960 Col=21 Row=D"  
/sex="male"  
/clone\_lib="RPCT-11 Human Male BAC Library"

## ORIGIN

/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"

Query Match 18.1%; Score 362.2; DB 9; Length 764;  
Best Local Similarity 82.7%; Pred. No. 1.4e-48;  
Matches 454; Conservative 0; Mismatches 83; Indels 12; Gaps 3;  
QY 517 AATGACCAATCAGCACTCTGTAATGACCAATCAGCAATGAGAGTGGAGGCTCAAA 576  
Db 647 AATGGACCAATTAGCACTCTGTAATGACCAATCAGCAATGAGAGTGGAGGCTCAAA 588  
QY 577 TAAAGGATGTAATGAGGACCGACCGACCGACCGACCGACCGACCGACCGACCG 632  
Db 587 TAAAGGATGTAATGAGGACCGACCGACCGACCGACCGACCGACCGACCGACCG 528  
QY 633 TCCACACTGAGAGGCTTTGTTGCTTTCAACAATAATCTTGTGCTGCTGATTC 692  
Db 527 CACCCCTTGGAGGCTTTGTTGCTTTCAACAATAATCTTGTGCTGCTGATTC 468  
QY 693 TTTGTGTCACACTACTTTATGAGCTGTAACTCACTGCGAGGCTGTGCTTCATT 752  
Db 467 ATTGGGCTCCCTACTACTTTATGAGCTGTAACTCACTGCGAGGCTGTGCTTCATT 408  
QY 753 CCTGAAGTCAAC-AGACACGAACCCACTGGAAGAAACAAAGAACTCCGATGCTGCC 811  
Db 407 CCTGAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTCCGAGCTGCACC 348  
QY 812 TTTAAGAGCTGTAACTCACTGCGAAGCTCTGCACTTCACTCTGTAAGTCACTGAGAC 871  
Db 347 TTTAAGAGCTGTAACTCACTGCGAAGCTCTGCACTTCACTCTGTAAGTCACTGAGAC 288  
QY 872 CACAAACCCACGAGAAAGAAACCTGTGACACACCTGTAATGTAAGAAACAACTC 931  
Db 287 CAGGAACCCACGAGAAAGAAACCTGTGACACACCTGTAATGTAAGAAACAACTC 228  
QY 932 CAGACACACCATCTTTGAGAGCTGTAACTCACTGCGAAGGCTGTGCTTCATTCTTG 991  
Db 227 CCGACACACCATCTTTAAGAGCTGTAACTGTA-----CGTAGGGCTGTGCTTCATTCTTA 175  
QY 992 AAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 1051  
Db 174 AAGTCAAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTGTGACACGATGATGACAA 115  
QY 1052 TATTTTGA 1060  
Db 114 AATGTTAGA 106

RESULT 32  
CB984929 732 bp mRNA linear EST 01-MAY-2003  
LOCUS AGENCOURT\_13647005 NIH\_MGC\_184 Homo sapiens cDNA clone  
DEFINITION IMAGE:30327682.5', mRNA sequence.  
CB984929  
ACCESSION CB984929  
VERSION CB984929.1 GI:30279453  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 732)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkevits  
cDNA Library Preparation: CLOUTEC Laboratories, Inc.

cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDCM137 row: b column: 11  
High quality sequence stop: 488.

## FEATURES

source

```
1. 732
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30327682"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1ib="NIH_MGC_184"
/notes="Organ: Pooled-glandular; Vector: pDNR-LIB; Site_1:
SfiI (ggccattatggc); Site_2: SfiI (ggccgctcgggc);
Library is oligo-dT primed and directionally cloned. cDNA
was prepared from a glandular pool of tissues from thyroid,
parathyroid, adrenal, cortex and pineal gland. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGGCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.38
kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
```

## ORIGIN

Query Match 17.9%; Score 357.2; DB 6; Length 732;  
Best Local Similarity 80.4%; Pred. No. 9.1e-48;  
Matches 500; Conservative 0; Mismatches 98; Indels 24; Gaps 6;

```
369 CAGCAGCTGGGGTCTCTCTTTTGAGAGGGGATTGAGAGTGAAGCAGCTGGGCTTTCG 428
      |||
25  CATTAAATTTTGGGAAATCTGTGTAATGTTACTGAGAGGTGAAGTCAGATGAGCTTTCG 84
      |||
429 GGTCAAGTGGGAGCTTGAAGAACTTTTGTCTAGCTAAAGATTGAAATGACCAATC 488
      |||
85  GGTGGATGGGAGCTTGAAGAACTTTTGTCTAGCTAAAGATTGAAATGACCAATC 144
      |||
489 AGCAGCTGTGTCTAGCTTAAAGATTGAAATGACCAATGACACTCTGTTAAATGAC 548
      |||
145 AGCAGCTGT-----AAATGGGCAATGACACTCTGTTAAATGACGAC 187
      |||
549 CAATGACGAGATGTGGGGGGGTCAATTAAGGAGTAAATGCGGACCCAGCCAGC 608
      |||
188 CAACCAATGAGGATGGGAGGAGCAATTAAGGAAATAAAGCTGCTCCCAAGCCAGC 247
      |||
609 AGTGGCAACCACTCGGGTCCCTTCCACACTGTGAAGCTTGTCTTGTGCTTTCAC 668
      |||
248 AGTGGCAACCGGCTCGGGTCCCTTCCACACTGTGAAGCTTGTCTTGTGCTTTCAC 307
      |||
669 AATTAATCTTGTCTGTCTCATTTCTTGTGTCCACACTGCTTATGAGCTGTGAACCTC 728
      |||
308 CATTAATCTTGTCTGTCTCATTTCTTGTGTGTCCACTTGAAGAGCTGTGAACCTC 367
      |||
729 ACTGCGAGGCTGTGTCTCATTTCTGAAAGTCA-ACAGACCAAGAACCACTGGAAGGA 787
      |||
368 ACTGCGAGGCTGTGTCTCATTTCTGAAAGTCAAGAGGAACCAACCAAGAGGA 427
      |||
788 ACAAGAAATCCCGATGTGTCTTGAAGCTGTGAACCTGCTGCAAGAGCTGTGAG 847
      |||
428 ACACACAACTCCGAGCGGCACTTTTAAGCGCTTAACCTGACGCAAGAGCTGTGAG 487
      |||
848 CTTCAGCTCTGAAGTCA---GTGAGACCAACCAACCAAGAGAGAAACT-CTGAGAC 903
      |||
488 CTTCAGCTCTGAAGTCAAGAGAAACCATGAAACCAACCGGAAAGAGAAATCTCTGAG 547
      |||
904 ACACCTGAATATCTGAA-AGAAACAACTCGAGACACCACTTTTTCAGAGCTGTGAACCT 962
      |||
548 ACATCTGAACATCTGAAGGAACCACTCTGACACACTTTCTTTTAAACTGTGAACCT 607
      |||
```

QY 963 CAC-CCGAGAGGTCTGTGAGCTT 963  
DB 608 CCTTCGAAAGGCTGTGAGCTT 629

RESULT 33  
AG132879 675 bp DNA linear GSS 04-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-145G04.F, genomic survey sequence.  
DEFINITION AG132879  
ACCESSION AG132879  
VERSION 1 GI:16662557  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euteria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Pan.

REFERENCE  
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Toto, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of library PTB  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 675)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Toto, Y., Watanabe, H. and Sakaki, Y.  
DIRECT SUBMISSION  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: [chimpes@sc.riken.go.jp](mailto:chimpes@sc.riken.go.jp), [URL:http://hgp.gsc.riken.go.jp/](mailto:URL:http://hgp.gsc.riken.go.jp/),  
Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the Rad process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI

## FEATURES

source

```
1. 675
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-145G04.F"
/sex="male"
/cell_type="lymphoblast"
/clone_1ib="PTB Chimpanzee Male BAC Library"
```

## ORIGIN

Query Match 17.8%; Score 355.4; DB 10; Length 675;  
Best Local Similarity 79.8%; Pred. No. 1.8e-47;  
Matches 499; Conservative 0; Mismatches 81; Indels 45; Gaps 5;

```
211 GAGCTCAATCAACTTCTGTGAGGAGCCCTGAGCCGCTGAGCCCTTCAATG 270
      |||
82  GAGCTCAATCAACTTCTGTGAGGAGCCCTGAGCCGCTGAGCCCTTCAATG 141
      |||
271 CCTAAAGAGCTCCCTCTGAGGAGCACTACAGGAGCCCTTCTTCACTTATCC 330
      |||
142 CCTAAAGAGTCCCTCTGAGGAGCACTACAGGAGCCCTTCTTCACTTATCC 201
      |||
331 AGCAGAAATAGCTACAGGAGTCAATG-CCAAATCCCAACAGAGTGGGAGTCTGTT 389
      |||
202 AGCAGAAATAGCTACAGGAGTCAATG-CCAAATCCCAACAGAGTGGGAGTCTGTT 261
      |||
390 TGGAGGGGAGTTGAGAGGTGAAGCCAGCTGAGC-TTCTGGTCAAGTGGGAGCTTGAAG 448
      |||
262 TAGAGGGGAGTTGAGAGGTGAAGCCAGCTGAGCTTCTGCACTTTCTGAGGAGCTTGAAG 321
      |||
449 AACTTTTGTCTAGCTAAAGATTGTAATGCAACCAATCAGCACTGTGTCTAGCTAA 508
      |||
```

Db 322 AACTTT-----TCTGTACGTAA 340  
 QY 509 AGCATTTGAATACCAATACAGACTCTGTGTAATAGACCAATACAGAGATGTGGCG 568  
 Db 341 AGGTTTGTAAACCAAGTACAGACACTGTGTAATAGACCAATCGGACAGATGGGCA 400  
 QY 569 GGGTCAATTAAGGAGTAATAAAGTGGCAACCGAGCAGAGTGGCAACCACTGGGTC 628  
 Db 401 GGGCAATTAAGGAGTAATAAGAGCTGGACACCTAGAGCATAGGCGACCCGTTGGGTC 460  
 QY 629 CCCTTCACACATGAGAGCTTTGTTCTTTGCTCTTTCACATTAATCTTGTGTC 688  
 Db 461 CCCTTCACAGTGGAGAGCTGTTCTTTCTTTCATCTTCAATTAATCTTGTGTC 520  
 QY 689 ATTCTTTGTGTCCACATCACTCTTTATGAGCTGTGAACATCACTGCGAGGCTGTGGCT 748  
 Db 521 ACTCTTTGCTTGAACATCACTCTTTATGAGCTGTGAAC-TCACCAAGAGCGGTGTGGTTC 579  
 QY 749 CATTCCTGAAGTGAAC-AGACCAAGAACCCACAGAGAGAAAGAACTCCGAGTGC 807  
 Db 580 CATTCCTGAAGTGAACACCCCAACCCACCTTGGAGAGACCAAAATCCAGAGTGC 639  
 QY 808 TGCCTTTAAGAGCTGTGAACACTCAC 832  
 Db 640 CACCTTTACCAAGCTGTGAACATGAC 664

RESULT 34  
 AL603245/c 636 bp mRNA linear EST 04-SEP-2003  
 LOCUS DKFZp686N032\_r1 686 (synonym: hicc3) Homo sapiens cDNA clone  
 DEFINITION DKFZp686N032\_5, mRNA sequence.  
 ACCESSION AL603245  
 VERSION AL603245.1 GI:1516751  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 TITLE 1 (bases 1 to 636)  
 JOURNAL Wambuit, R., Heubner, D., Mewes, W., Weil, B. and Wiemann, S.  
 EST (Wambuit, R., Heubner, D., Mewes, H.W., Weil, B. and Wiemann, S.)  
 COMMENT Unpublished (1999)  
 CONTACT: MIPS  
 MIPS Ingolstaedter Landstr.1, D-85764 Neuberg, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by AGOMA (Berlin/Germany) within the cDNA sequencing  
 consortium of the German Genome Project.  
 No si sequence available.  
 This clone (DKFZp686N032) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 Location/Qualifiers  
 1..636  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp686N032"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="686 (synonym: hicc3)"  
 /note="Vector: pTriplEx2; site\_1: stf1a; site\_2: stf1b;  
 cDNA-collection"

ORIGIN  
 Query Match 17.7%; Score 354.8; DB 1; Length 636;  
 Best Local Similarity 86.3%; Pred. No. 2.3e-47;  
 Matches 428; Conservative 0; Mismatches 62; Indels 6; Gaps 3;

QY 564 GGGCGGGGTCAATAGGAGTAATAAACTGGCCACCCGAGCCAGCAGTGGCAACCCACTC 623  
 Db 636 GGGTGGGGCCAAATTAAGGAGTAATAAAAGAGGCCACCAAGCCAGCAGCGCAACTGCTT 577  
 QY 624 GGGTCCCTTCCACACTGTGAAGCTTTGTTCTTTGCTCTTTCACATTAATCTTGTCTGC 683  
 Db 576 GGGTCCCTTCCACAGCTGTGAAGCTTTGTTCTTTGCTCTTTCACATTAATCTTGTCTGC 517  
 QY 684 TGTCTATTCTTTGTGTCCACTACTCTTTATGAGCTGTGAACACTGCGAGGCTGTGT 743  
 Db 516 TGTCTACTCTTTGGGTCGCCACTACTCTTTATGAGCTGTAA-----CACTCGAAGGCTGTGC 461  
 QY 744 GGTCTATTCTCGAAGTGA-ACAGACCAAGAACCCACTGGAAGGAACAAGAACTCCCA 802  
 Db 460 AGCTTACTCTTAAAGCCAGTGAACCAAGAACCCCAAGAGGAACCAACCACTCCGA 401  
 QY 803 TGTGCTGCTTTAAGAGCTGTGAACACTCACTGCGAAGCTCTGAGCTTCACTCTGAAGT 862  
 Db 400 TGCACCACTTTAAGAGCTGTGAACACTCACTGCGAAGCTCTGAGCTTCACTCTGAAGT 341  
 QY 863 CAGTGAAGCCACCAACCCACAGAA-AGAAAGAACTTGAACACACTGTGAATCTGAAG 921  
 Db 340 CAGTGAAGCCATGAAGCCACCAAGAGGAAGAACTCCGACACATCTGAATCTGAAG 281  
 QY 922 GAACAACCTCCAGACACACATCTTTAGAGCTGTGAACACTCAAGCGCAAGGCTGTGCG 981  
 Db 280 GAACAACCTCCAGACACACATCTTTAAGAGCTGTGAACACTCAAGCGCAAGGCTCCCGCG 221  
 QY 982 TTCATTCTTGAAGTGAAGACCAAGAACCCAGAGGAAGAAATTCAGACACACT 1041  
 Db 220 TTCATTCTTGAAGTGAAGACCAAGAACCCAGAGGAAGAAATTCAGACACT 161  
 QY 1042 AGCAATCTGTATTTT 1057  
 Db 160 AGCATATTCGCACTTT 145

RESULT 35  
 AQ747612/c 845 bp DNA linear GSS 19-JUL-1999  
 LOCUS HS 5533 A2 C11 T7A RPCI-11 Human Male BAC library Homo sapiens  
 DEFINITION genomic clone Plate=1109 Col=22 Row=E, genomic survey sequence.  
 ACCESSION AQ747612  
 VERSION AQ747612.1 GI:5534770  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 845)  
 TITLE Mahiwas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
 Hood, L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 JOURNAL 10449764  
 COMMENT High Throughput Sequencing Center  
 Contact: Mahiwas G.G., Wallace J.C., Hood L.  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BAC PAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
 or from Research Genetics (info@resgen.com). BAC end Web Server:  
 http://www.htsc.washington.edu  
 Plate: 1109 row: E column: 22



Oy	833	GCTCTGACGTTTCATCTCTGTAAGTGAAGAGACCAAAACCCACCAAGAAAGAAATC	833
Db	157	GGTCTGCGGCTTATCTCTGAAGTGAAGACCAAGAAACCACAGAAAGAAATC	98
Oy	899	TGACACACCTGAAATCTGGAAGAAACAACTCCGACACACATCTTTCAGACTTAA	958
Db	97	TAAACACATCTGAACATCTGGAAGAAACAACTTGAACGACCATTTTAAGAACTTAA	38
Oy	959	CACCTACCGCAAGGCTGTGTGGCTTCATTTTGAAGT	995
Db	37	CAATCACTTGAAGGCTCTCGGCTTCATTTGAATT	1
RESULT 38			
LOCUS	AAQ239957/c	680 bp	DNA linear GSS 30-SEP-1999
DEFINITION	CIT-HSP-2387K11.FR.1 CIT-HSP Homo sapiens genomic clone 2387K11,		
ACCESSION	AAQ239957		
VERSION	AAQ239957.1	GI:3672155	
KEYWORDS	GSS.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 680) Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.		
TITLE	Use of a random human BAC End Sequence Database for Sequence-Ready Map Building		
JOURNAL	Unpublished (1998)		
COMMENT	Other GSSs: CIT-HSP-2387K11.FR.1 Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hunggen/bac_end_search/bac_end_search.html. Seq primer: M13-21 Class: BAC ends.		
FEATURES			
source	Location/Qualifiers		
	1..680		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/clone="2387K11"		
	/sex="Male"		
	/cell_type="Sperm"		
	/clone_1ib="CIT-HSP"		
	/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"		
ORIGIN			
Query Match	17.1%;	Score 341.4;	DB 9; Length 680;
Best Local Similarity	84.0%;	Pred. No. 3.3e-45;	
Matches 421; Conservative	0;	Mismatches 76;	Indels 4; Gaps 3
Oy	517	AAATGACCAATCAGCACTGTGTAATAATGACCAATCAGCAGATGTGGCGGGGTCAA	576
Db	629	AAACACACCAATGGGCTCTCTGTAAATGACCAATCAGCAGATTTGGGTGGGGCCAGA	570
Oy	577	TAAAGGAGTAATAAACTGGGCACCGACGACAGTGGCAACCCACTGGGTCCCTTCCA	636
Db	569	TAAAGAAATAAAGCAGGGCTGCCAGAGCCAGACGTGGCAACCTCTTGGGTTCCCTTCCA	510
Oy	637	CACCTGGAAGCTTGTGTTCTTTTGCTCTTCACAATAATCTTGCTGCTGCTCATTTCTTG	696

```

Db      509  CACTGTGGAGGCTTTTGTCTTTGCTCTTACATATAATCTTCTGCTGCTACCTCTTG 450
QY      697  TGTCCACACTACTTATGAGTGTAACTCACTCGAGGGTCTGTGCTTCAATCTTG 756
Db      449  GGTCCACACTACTTATGAGTGTAACTCACTCGAGGGTCTGTGCTTCAATCTTG 390
QY      757  AAGTCAAC-AGACCAGAACCCACTGGAAGAACAAGAACTCCGATGTGCTGCTTGA 815
Db      389  AAGCCAGCAAGACCATGAAACCATCAGAGAGATGAACAACCTCAGATGTG--ACTTAA 332
QY      816  AGAGCTGTAACTACTGTGGAAGCTCTGTGAGCTTCACTCTGAAGTCAAGTGAACACA 875
Db      331  AGAGCTGTAACTACTGTGGAAGCTCTGTGAGCTTCACTCTGAAGTCAAGTGAACACA 273
QY      876  AACCCACGAAAGAAAGAACTCTGAGACACACCTGAATATCTGAAGAAACAATCTCGA 935
Db      272  AACCCACGAAAGAAAGAACTCTGAGACACACCTGAATATCTGAAGAAACAATCTCGA 213
QY      936  CACACCATCTTTAGAGCTGTAACTCACTCAGCGAAGGCTGTGCTTCAATCTTGAAGT 995
Db      212  CATGTGCTTTAAGAACTGTAACTCACTCACTGCGAGGGTCCGGGCTTCACTTAAAGT 153
QY      996  CAGCAGACCAAAACCCACC 1016
Db      152  CAGTGAGACCAAGAACTCAGC 132

RESULT 39
B66172/c 566 bp DNA linear GSS 21-JUN-1998
LOCUS     B66172
DEFINITION CIT-HSP-2024E20.TR CIT-HSP Homo sapiens genomic clone 2024E20,
           genomic survey sequence.
ACCESSION B66172
VERSION   B66172.1 GI:2640150
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homiidae; Homo.
REFERENCE 1 (bases 1 to 566)
AUTHORS  Adams,M.D., Rounsley,S.D., Field,C.E., Baas,S., Linher,K.,
         Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
         Simon,M. and Venter,J.C.
         Use of a random BAC End Sequence Database for Sequence-Ready Map
         Building
JOURNAL   Unpublished (1997)
COMMENT   Other-GSSs: CIT-HSP-2024E20.TF
           Contact: Mark Adams
           Department of Bukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: mdamadams@tigr.org
           Clones are available from Research Genetics (info@resgen.com). BAC
           end search page:
           http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html
           Seq primer: M13 Reverse
           Class: BAC ends.
FEATURES
   source          1..566
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="GDB:7046518"
                    /db_xref="taxon:9606"
                    /clone="2024E20"
                    /sex="Male"
                    /cell_type="Sperm"
                    /clone_1db="CIT-HSP"
                    /note="Vector: pBelBAC11, Site_1: HindIII, Site_2:
                    HindIII"

```

```

ORIGIN
Query Match      17.0%; Score 340.6; DB 9; Length 566;
Best Local Similarity 81.1%; Pred. No. 4,7e-45;
Matches 455; Conservative 0; Mismatches 54; Indels 52; Gaps 3;

QY      141  TGTCAAAATTTTCTCTCTAGAGATCGAGCCATCAAGTACAGATGATCTTACAAATCT 200
Db      565  TGTCAAAATTTTCTCTCTAAATCAAGGCAATCAAGTACAGATGATCTTACAAATCT 506
QY      201  AACCCCAATGAGTCAACTAACACTTCTGTGAGAGACCCCTGAGACGACCCGCTGACC 260
Db      505  AACCCCAATGAGTCAACTAACACTTCTGTGAGAGACCCCTGAGACGACCCGCTGACC 446
QY      261  CTTTCAATGGCTTAAAGAGCTCCCTCTGAGAGACATCACTCAGGGCCCTCTTCTTTC 320
Db      445  ATTTCACATGGCTTAAAGAGTTTCCCTCTGAGAGACACCACTCAGGGCCCTCTTCTTTC 386
QY      321  ACCCTTATCAGACAGAAAGTACTACAGCGGTCAATCG--CCAAATCCCAACAGCAGCTTG 378
Db      385  ACCCTTATCAGACAGAAAGTACTACAGCGGTCAATCGAAATTCCAAAACAGCAGTTGG 326
QY      379  GGTGTCCTGTGAGAGGGGAGATTGAGAGTGAAGCCAGCTGGGCTTCTGGGTCAAGTGG 438
Db      325  GGTGTCCTGTGAGAGGGGAAATTGAGAGTGAAGCCAGCTGGGCTTCTGGGTGGGTGG 266
QY      439  GGAATTTGAGAACTTTTGTCTAGCTTAAAGGATTTGTAATGACCAATCAGCACTCTGT 498
Db      265  GGAATTTGAGAACTTTTGTCTAGCTTAAAGGATTTGTAATGACCAATCAGCTCTGT 206
QY      499  GTCTAGCTTAAAGATTTGTAATGACCAATCAGCA----- 533
Db      205  GTCTAGCTTAAAGATTTGTAATGACCAATCAGCACTTAAATGAGCAATCAGCA 146
QY      534  -----CTCTGTAATATGACCAATGACCAATGAGAGATGTGGGCGG 569
Db      145  CTCTGTAATATGACCAATCAGCTCTCTGCAAAATGAGATCTCAAGTAGATGCAAGTGG 86
QY      570  GGTCAAAATTAAGGAGTAAATTAAGTGGCACTGAGCCAGCGCAGGATGGCAACCACTCGGGTCC 629
Db      85  GGTCAAAATTAAGGAGTAAATTAAGGCACTGAGCCAGCGCAGGATGGCAACCACTCGGGTCC 26
QY      630  CCTTCCACACTGTGGAAGCTT 650
Db      25  CC-TCCACACAGTGAAGCTT 6

RESULT 40
AQ481794/c 497 bp DNA linear GSS 24-APR-1999
LOCUS     AQ481794
DEFINITION RPCI-11-237L13.TV RPCI-11 Homo sapiens genomic clone
ACCESSION AQ481794
VERSION   AQ481794.1 GI:4669198
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homiidae; Homo.
REFERENCE 1 (bases 1 to 497)
AUTHORS  Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and
         Venter,J.C.
         Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
         Map Building
JOURNAL   Unpublished (1997)
COMMENT   Other-GSSs: RPCI-11-237L13.TV
           Contact: Shaying Zhao, William Niernan, Mark Adams
           Department of Bukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850
           Tel: 301 838 0200
           Fax: 301 838 0208

```



Email: hbeef@r.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..497

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="GDB:759090"

/db\_xref="taxon:9606"

/clone="RPCI-11-237L13"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; RPCI11 Human Male Library"

## ORIGIN

Query Match 17.0%; Score 340.4; DB 9; Length 497;  
Best Local Similarity 83.2%; Pred. No. 5.2e-45;  
Matches 411; Conservative 0; Mismatches 81; Indels 2; Gaps 2;

523 ACCAATCAGCAGCTCTGTAATAATGACCAATCAGCAGATGTGGCGGGTCAATAAGG 582  
|||||  
493 ACCAATCAGCAGCTCTGTAATAATGACCAATCAGCAGATGTGGCGGGTCAATAAGG 434  
|||||  
583 AGTAAAACTGGCCACCGGACGACGATGGCAACCACTCGGCTCCCTTCCACTGT 642  
|||||  
433 AGTAAAAAGAGGCTGCGCCAGCAGCACTGGCAACCTTCGGGTCGGTTCATCTGT 374  
|||||  
643 GGAAGCTTGTGTTTGTCTTCTTCAATTAATCTTGTGCTGCTCATTTCTTGTGCTCA 702  
|||||  
373 GGAAGCTTGTGTTTGTCTTCTTCAATTAATCTTGTGCTGCTCATTTCTTGTGCTCA 314  
|||||  
703 CACTACCTTATAGAGCTGTAACTCACTGCGAGGCTGTGCTTCTTCTTCTTGAAGTCA 762  
|||||  
313 CGTGCTTATAGAGCTGTAACTCACTGCGAGGCTGTGCTTCTTCTTCTTGAAGTCA 254  
|||||  
763 AC-AGACGAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 821  
|||||  
253 GCGAGACCAAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 195  
|||||  
822 GTAACTCACTGCGAGGCTGTGAGCTTCACTGCGAGGCTGTGAGCTTCACTGAGTCA 881  
|||||  
194 GTAACTCACTGCGAGGCTGTGAGCTTCACTGCGAGGCTGTGAGCTTCACTGAGTCA 135  
|||||  
882 CCAGAGAGAAAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 941  
|||||  
134 CCAGAGAGAAAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 75  
|||||  
942 ATCTTCAAGCTGTAACTCACTGCGAGGCTGTGAGCTTCACTGAGTCAAGCA 1001  
|||||  
74 GCTTTAAGAACTGTAACTCACTGCGAGGCTGTGAGCTTCACTGAGTCAAGTGA 15  
|||||  
QY 1002 GACCAAGAAAGCCAC 1015  
|||||  
DB 14 GACCAAGAAAGCCAC 1

## RESULT 41

BE393102 666 bp mRNA linear EST 21-JUL-2000  
LOCUS BE393102  
DEFINITION 601309151F1 NIH\_MGC\_44 Homo sapiens cDNA clone IMAGE:3630709 5',  
mRNA sequence.  
ACCESSION BE393102  
VERSION BE393102.1 GI:9338467  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

## REFERENCE

1 (bases 1 to 666)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapdb-r@mail.nih.gov](mailto:cgapdb-r@mail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LCM319 row: e column: 14

High quality sequence stop: 609.

Location/Qualifiers

1..666

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3630709"

/issue\_type="endometrium, adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 16.9%; Score 338.4; DB 2; Length 666;  
Best Local Similarity 83.9%; Pred. No. 1e-44;  
Matches 418; Conservative 0; Mismatches 76; Indels 4; Gaps 3;

520 TGACCAATCAGCAGCTCTGTAATAATGACCAATCAGCAGATGTGGCGGGTCAATAAG 579  
|||||  
1 TGACCAATCAGCAGCTCTGTAATAATGACCAATCAGCAGATGTGGCGGGTCAATAAG 60  
|||||  
580 GGGAGTAAAACTGGCCACCGGACGACGATGGCAACCACTCGGCTCCCTTCCACAC 639  
|||||  
61 GGAATTAAGAGAGGCTGCGAGTAAAGCAGTGGCAACCGGCTCGGCTCACTTCGACAC 120  
|||||  
640 TGTGGAAGCTTGTGTTTGTCTTCTTCAATTAATCTTGTGCTGCTCATTTCTTGTGT 699  
|||||  
121 TGTGGAAGCTTGTGTTTGTCTTCTTCAATTAATCTTGTGCTGCTCATTTCTTGTGT 180  
|||||  
700 CCACACTTCTTATAGAGCTGTAACTCACTGCGAGGCTGTGAGCTTCACTGAGTCA 759  
|||||  
181 CCACACTTCTTATAGAGCTGTAACTCACTGCGAGGCTGTGAGCTTCACTGAGTCA 240  
|||||  
760 TCAAC-AGACCAAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 818  
|||||  
241 CCATGAGACCAAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 299  
|||||  
819 GCTGTAACTCACTGCGAGGCTGTGAGCTTCACTGCGAGGCTGTGAGCTTCACTGAGTCA 878  
|||||  
300 GCTGTAACTCACTGCGAGGCTGTGAGCTTCACTGCGAGGCTGTGAGCTTCACTGAGTCA 357  
|||||  
879 CCACGAGAGAAAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 938  
|||||  
358 CCACGAGAGAAAGAACTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 417  
|||||  
939 ACCATTTTCAAGCTGTAACTCACTGCGAGGCTGTGAGCTTCACTGAGTCAAG 998  
|||||  
418 ACCGCTTTAAGAACTGTAACTCACTGCGAGGCTGTGAGCTTCACTGAGTCAAG 477

Yr	999	CAAGACCAAGAACCCACC	1016	
Db	478	TGAAGCCAGAACCCACC	495	
RESULT 42				
AI276815/c				
LOCUS				
DEFINITION	AI276815	424 bp	mRNA	linear EST 28-JAN-1999
ACCESSION	gk1d07.x1	NCI-CGAP Co8 Homo sapiens cDNA clone IMAGE:1873453	3'	
VERSION	AI276815			
KEYWORDS				
SOURCE	AI276815.1	GI:3899089		
ORGANISM	EST.			
	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1	(bases 1 to 424)		
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
JOURNAL	Unpublished (1997)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgsb@remail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the 1.M.A.G.E. Consortium/LNLW at: www-bio.lnl.gov/bbip/image/image.html Insert Length: 873 Std Error: 0.00 Seg primer: -40UP from Gibco High quality sequence stop: 466. Location/Qualifiers			
FEATURES				
source	1..424			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:1873453"			
	/tissue_type="adenocarcinoma"			
	/lab_host="DH10B"			
	/clone_idb="NCI CGAP Co8"			
	/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - cIio4(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."			
ORIGIN				
Query Match	16.9%;	Score 338.2;	DB 1;	Length 424;
Best Local Similarity	90.2%;	Pred. No. 1.2e-44;		
Matches	385;	Conservative 0;	Mismatches 38;	Indels 4;
				Gaps 2;
Yr	81	TAGGTCCTGTGACGACCATCTTGCTAATATGTCGATTTGGGCGCTGTATTTTAACTCT	140	
Db	424	TAGGTCCTGTGACGACCATCTTGCTAATATGTCGATTTGGGCGCTGTATTTTAACTCT	366	
Yr	141	TGGTCAATTTGTTTCCCTCTAGATGCGAGGCGCATCAAGCTACAGTATCTTACAAATGT	200	
Db	365	--GTCAATTTGTTTCCCTCTAGATGCGAGGCGCATCAAGCTACAGTATCTTACAAATGG	308	
Yr	201	AACCCCAATGAGCTCAACTAACAATCTTGCTGAGAGACCCCTGAGCCGAGCCGCTGGCC	260	
Db	307	AACCCCAATGAGCTCAACTAACAATCTTGCTGAGAGACCCCTGAGCCGAGCCGCTGGCC	248	
Yr	261	CTTTCATAGGCTTAAGAGCTCCCTCTGAGAGACATCAACATGCGAGGCGCCCTTCTTC	320	

Db	247	CTTTCATGCGCTAAAGATTCCCTCCGAGGACACTAACAATCGAAGCCCTTCTTC	188
QY	321	ACCCCTATCCAGCAGAACTAGTACTACAGCGGTCATCG-CCAAATCCCAACAGACGCTGGG	379
Db	187	GCCCCCATCCAGCAGGAATGAGCCACAGCATCATATGCCCAATTCGCCMACAGCAGATTGGG	128
QY	380	GTCGCCGTTTGGAGGGGGGATTTGAGAGGATGAAGCCAGCGCTGGCTTCGGGTCAAGTGGG	439
Db	127	GTGTCCTGTTTAAAGGGGGCATTGAGAGGTAAGCCAGCTGGGCTTTCGGGTGGGTTGGG	68
QY	440	GACTTGAGAACTTTTGTCTAGTCTAAAGATTTGTAATGACCAATCAGACTCTGTG	499
Db	67	GACTTAAGAACTTTTGTCTAGTCTAAAGATTTGTAACACACCAATGAGCACTCTGTG	8
QY	500	TCTAGCT 506	
Db	7	TC TAGTT 1	
RESULT 43			
LOCUS	AG113408/c	681 bp	DNA linear GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-120A12.R, genomic survey sequence.		
ACCESSION	AG113408		
VERSION	AG113408.1 GI:16733927		
KEYWORDS	GSS.		
SOURCE	Pan troglodytes (chimpanzee)		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae; Pan.		
AUTHORS	1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	BAC end sequences of library PTB		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 681)		
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-AUG-2001) Ageo Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)		
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.		
PRIMERS			
Sequencing: M13Rev			
LIBRARY			
Vector	: pKS145		
R.Site 1	: SacI		
R.Site 2	: SacI.		
Location/Qualifiers			
1..681			
/organism="Pan troglodytes"			
/mol_type="genomic DNA"			
/db_xref="taxon:9598"			
/clone="PTB-120A12.R"			
/sex="male"			
/cell_type="lymphoblast"			
/clone_lib="PTB Chimpanzee Male BAC Library"			
ORIGIN			
Query Match	16.9%;	Score 337.4;	DB 10; Length 681;
Best Local Similarity	78.3%;	Pred. No. 1.5e-44;	
Matches 474;	Conservative 0;	Mismatches 86;	Indels 45; Gaps 4;
QY	452	TTTGTGCTAGCGAAGAGATTGTAATGACCAATCAGCAGCTGTGTGT-----CTA	503
Db	649	TTCTGTGCTAGCTAAAGCTTTGTGAATGTATGCCAATCAATGACTCTGTAAAAATGACACCA	590

OY		504	GCTAAGAAGATTGTAATGACCAACAATCAGACACTGTGTAATAATGGACAATTCAGACAGAAAGT	563
Db		589	TGAGACACTCTGTAAAAACGACCMAATCAGACACTGTGTAATAAGGGACTTAATCAGACAGAAAGT	530
OY		564	GGCGGGGGGTCAATATAGGAGAGTAAAAAACTGGCCACCCGACCCGACGAGTGGCACACCCTC	623
Db		529	GAGTGGGGGCCAAATATAGGGAATPAAAAGACAGGCCAACCAAGCCAGACGCAACCCCTCTT	470
OY		624	GGGTCCCCCTTCCACACTGTGGAAAGCTTTGTTCTTTGCTCTTCACAATTAATCTTGCTGC	683
Db		469	GGGTCCCCTTCCACGCTGTGGAAAGCTTTGTTGTTTGTCTTTAGAAATTAATCTTGCTGC	410
OY		684	TGCTCATTTCTTTGTGTCCACACTACTCTTTATAGCTGTAACTCACTCACTCGAGGGTCTGT	743
Db		409	TGCTCACTAGTTAAAGTCCACACTTAATTATGTGTCTGTAACTCACTCACTCGAGGGTCTGC	350
OY		744	GGCTTCATTCTCTGAAGTCAAC-AGAACAAGAACCCACTGGAAGGAACAAGAATCTCCGA	802
Db		349	AGCTTCACCTCCGTAAGTCAAGAAAGTATGAACCCACGAGGAGTAATGAAGCACTCCGGA	290
OY		803	TGTGTGCTCTTTAAAGCTGTAACTCACTCACTCGAAAGCTCTGCAGCTTCACTCCTGAAGT	862
Db		289	CCTGCACACTTTTAAAGCTGTAACTCACTCACTGTGAAGGTCTGCAGCTTCACTCCTGAAGT	230
OY		863	CAGTGAGACCAACAACCCACGAAAGGAAGAAACTGTGGAACAACCTGAATATGTGAAGG	922
Db		229	CAGCAAGACTACAAAC-----ATCTGAAGG	205
OY		923	AACAAACTTCACAGC-ACACCATCTTTCAAGCTGTAACTCACTCACTCGAAGGGTCTGTGdC	981
Db		204	AACAAACTCTGGACGACACATCTTTAAGAACTGTAAACACTCACTCACTGAAGGGTCCGGdC	145
OY		982	TTTCATTCTTGAAGTAGCAAGAACCAGAAACCACCGAAGGAACAATTCCAGACACAGT	10411
Db		144	TTTCATTCTTGAATCAGCCAGCACCAAGAACCCACTGGAAGGAACCAATTCCAGACACAYT	85
OY		1042	AGGAA	1046
Db		84	GCGAA	80

RESULT 44  
 A0083508 LOCUS 600 bp DNA linear GSS 20-APR-1999  
 DEFINITION RPII11-52J19..TJ RPII-11 Homo sapiens genomic clone RPII-11-52J19,  
 genomic survey sequence.  
 A0083508  
 ACCESSION A0083508.1 GI:344692  
 VERSION  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 600)  
 Adams,M.D., Rounsley,S.D., Zhao,S., Base,S., Linher,K., Golden,K.,  
 Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.  
 Use of human BAC End Sequences for Sequence-Ready Map Building  
 Unpublished (1998)  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel.: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are derived from the human BAC library RPII-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering/>) or from  
 Research Genetics ([incorresgen.com](http://incorresgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html)

FEATURES	Class: BAC ends.	Location/Qualifiers
source	1..600	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="GDB:7519818" /db_xref="taxon:9606" /clone="RPCL11-52J19" /sex="Male" /cell_type="Lymphocytes" /clone_id="RPCL11" /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; RPCL11 Human Male BAC Library"
Query Match	16.7%; Score 333.8; DB 9; Length 600;	
Best Local Similarity	77.7%; Pred. No. 5.7e-44;	
Matches 439; Conservative	0; Mismatches 77; Indels 49; Gaps 1;	
QY	380 GTGTCCTGTTTGGAGGGGGGATTAGAGGTGACACCGCTGGCTTTGGGTACGGTGGG	439
Db	32 GTACCTTAGAGCTGAGTGTCTGTAGAGGTGAACCTGTGGGCTCCTGGGTCCGGTGGG	91
QY	440 GACTTGGAGAACTTTGTGTACTGTAAAGGATGTAAATGACCAATTCAGCACTGTGTG	499
Db	92 GCTTGGAGAACTTTTGTGTACTGTAAAGGATGTAAATGACCAATTCAGCGCTGTGTG	151
QY	500 TCTAGCTAAAGATTTGT-----	516
Db	152 TCTAGCTAAATGTTTGTAAACGACCAATCAGCACTCTGTAAACACAGACTGATCAGCAC	211
QY	517 -----AATGTCACCAATCAGCACTCTGTAAATGACCAATTCAGCAGATGTGGGGGG	570
Db	212 TCTGTAAATGACCAATCAGCACTCTGTAAATGACCAATTCAGCAGAAATGAGTGGG	271
QY	571 GTCAATAATGAGAGTAAAACTGGCCACCCGAGCCAGCACTGGCACTCGGGTCCG	630
Db	272 GCCAAATTAAGGAAATTAAGCTGGCCACCCAGCCAGCAGCGGCAACCCGCTCAGGTCC	331
QY	631 CTTCCACACTGTGAGAGCTTTGTCTTTTGTCTTCAATTAATCTTGTGCTGCTCAT	690
Db	332 CTTCTATGCTGTGAAAGCTTTGTCTTTTGTCTTCAATTAATCTTGTGCTGCTCAT	391
QY	691 TCTTTGTGTCCACACTCTTTATGAGCTGTAACTCAGCTGCGAGGGTCTGTGGCTTCA	750
Db	392 TCTTTGGGGCCACACTACCTTTATGAGCTGTAACTCAGCTGCGAAGTCTGTGCTTCA	451
QY	751 TTCTGTAAATCAAGACCAAGAACTGGAAGAAACAAAGATTCCTGGATGTGCTGC	810
Db	452 TCTGGAAGTCAAGCAATGAACCCACCGGGGAGAAAGTAACTCCGGATCCACAC	511
QY	811 CTTTAAGAGCTGTAACTCAGCTGCGAAGCTGCGACCTTCACTCCGTAAGTCAAGTGA	870
Db	512 CTTTATAGCTGTAACTCAGCTGCGAAGGCTGCGACCTTCACTCCTGAAGTCAAGTGA	571
QY	871 CCACAAACCAACCAAGAGAA 895	
Db	572 CCATGAACCAACCAAGAGGACAA 596	
RESULT 45		
AQ507503/c	556 bp DNA linear GSS 29-APR-1999	
LOCUS	RPCL11-298B4.TU RPCL11 Homo sapiens genomic clone RPCL11-298B4,	
DEFINITION	genomic survey sequence.	
ACCESSION	AQ507503	
VERSION	AQ507503.1 GI:4712250	
KEYWORDS	GSS.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;	
	Hominidae; Homo.	

REFERENCE 1 (bases 1 to 556)  
AUTHORS Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter,J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other GSSs: RPCI-11-298B4.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tifg.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([inforesgen.com](http://inforesgen.com)). BAC end search page: [http://www.tifg.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tifg.org/tldb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source Location/Qualifiers

1..556  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7614075"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-298B4"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; RPC11 Human Male BAC Library"

ORIGIN

Query Match 16.7%; Score 333.2; DB 9; Length 556;  
Best Local Similarity 87.4%; Pred. No. 7.3e-44;  
Matches 376; Conservative 0; Mismatches 53; Indels 1; Gaps 1;

```
QY 614 CAACCCACCTCGGGTCCCTCCCACTGTGGAGCTTTGTTCTTTGCTTTTACAAATAA 673
   |||
DB 556 CAACCCACCTCGGGTCCCTCCCACTGTGGAGCTTTGTTCTTTGCTTTTACAAATAA 497
   |||
QY 674 ATCTTGTGCTGCTCAATCTTTGTGTCACTACCTTTATGAGCTGTAACTCACTGC 733
   |||
DB 496 ATCTTGTGCTGCTCACTCTTTGTGTCACTACCTTTATGAGCTGTAACTCACTGC 437
   |||
QY 734 GAGGGTCTGTGGCTTCAATTCCTGAAGTCAAC-AGACCAGAACCACTGGAGAA 792
   |||
DB 436 GAGGGTCCGGTTCATTCCTCAAGTCAGCGAGATCAAAACCAAGAGAGCAAA 377
   |||
QY 793 GAACTCCCGATGTGCTGCTTTAAGAGCTGTAACTCACTGGAAGCTTGCAGCTTCA 852
   |||
DB 376 TAACTCCAGACACACACCTTTAAGAGCTGTAACTCACTGGAAGCTTGCAGCTTCA 317
   |||
QY 853 CTCTGAAGTCACTGAGACACAAACCAACCAAGAGAACTTGGACACACCTGAA 912
   |||
DB 316 CTCTGAAGTCACTGAGACACAAACCAACCAAGAGAACTTGGACACACCTGAA 257
   |||
QY 913 TATCTGAAGAAACAACTCCAGACACACATCTTTCAAGCTGTAACTCACTGCAAG 972
   |||
DB 256 CATCTGAAGAAACAACTTTAAGAGCTGTAACTTTAAGAGCTGTAACTCACTGCAAG 197
   |||
QY 973 GTCTGTGGCTTCAATCTTTGAAGTCAGAGACCAAGACCAACCGAAGAGAAATAATTC 1032
   |||
DB 196 GTCCAGGGCTTCACTCCCTGAAGTCAGACCAAGAACCAACCAAGAGACTAAATTC 137
   |||
QY 1033 AGACACAGTA 1042
   |||
DB 136 GGAACAATA 127
```

Search completed: January 22, 2006, 21:09:21  
Job time : 5433.54 secs



92	687.2	34.3	164486	8	CNS05TES	ALJ59332	Human chr
C	93	684.4	34.2	106938	8	AC132032	Homo sapi
C	94	684.2	34.2	165090	8	HS44A20	AL035086
C	95	684	34.2	144665	8	CNS07BE7	Human chr
C	96	683.8	34.2	71032	14	AC084084	Homo sapi
C	97	683.4	34.2	103188	8	AL731551	Human DNA
C	98	682.4	34.1	94219	8	AY796203	Homo sapi
C	99	682.4	34.1	115177	8	AL583836	Human DNA
C	100	682.2	34.1	130724	8	AC004253	Homo sapi

## ALIGNMENTS

RESULT 1  
AX033938  
LOCUS AX033938 39678 bp DNA linear PAT 21-SEP-2000  
DEFINITION Sequence 30 from Patent WO9851790.  
ACCESSION AX033938  
VERSION AX033938.1 GI:10280507  
KEYWORDS  
SOURCE unidentifed  
ORGANISM unidentifed  
REFERENCE 1 unclassified sequences.  
AUTHORS Cancelli,M.R., Choo,K.H. and Du,S.D.  
TITLE A novel nucleic acid molecule  
JOURNAL Patent: WO 9851790-A 30 19-NOV-1998;  
CANCELLI MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART  
DESIRE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)  
LOCATION/Qualifiers  
FEATURES  
source 1..39678  
/organism="unidentifed"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 100.0%; Score 2001; DB 6; Length 39678;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAGGCTTGAATATGACATATGCTTCAAACTCTTATACCAACTCTGAGTTGGG 60
DB 18083 AAAGGCTTGAATATGACATATGCTTCAAACTCTTATACCAACTCTGAGTTGGG 18142
QY 61 GACATGCTTCTCCCTTTCTAGGTCCTTGACAGCCATCTTGTAATAGTCGATTGG 120
DB 18143 GACATGCTTCTCCCTTTCTAGGTCCTTGACAGCCATCTTGTAATAGTCGATTGG 18202
QY 121 GCCCTGATTTTAACTCTTGTGCAATTTGTTCCCTAGAGTCGAGGCAATCAAGT 180
DB 18203 GCCCTGATTTTAACTCTTGTGCAATTTGTTCCCTAGAGTCGAGGCAATCAAGT 18262
QY 181 ACAGATGATCTTACAAATGTAAACCCCAATGAGCTCAACTAACACTTCTGCTGAGACC 240
DB 18263 ACAGATGATCTTACAAATGTAAACCCCAATGAGCTCAACTAACACTTCTGCTGAGACC 18322
QY 241 CTTGACCGACCGCGCTGACCTTTCAATGAGCTTAAAGAGCTCCCTCTGAGGACACTAC 300
DB 18323 CTTGACCGACCGCGCTGACCTTTCAATGAGCTTAAAGAGCTCCCTCTGAGGACACTAC 18382
QY 301 CACGCGAGGGCGCCCTTCTTCAACCCCTATCCAGCAGGAAGTACTACAGCGGTCATCGCA 360
DB 18383 CACGCGAGGGCGCCCTTCTTCAACCCCTATCCAGCAGGAAGTACTACAGCGGTCATCGCA 18442
QY 361 AATCCCAACAGCACTGGGTGTCTCTGTTTGAGAGGGGATTTGAGGTGAGCAAGCTG 420
DB 18443 AATCCCAACAGCACTGGGTGTCTCTGTTTGAGAGGGGATTTGAGGTGAGCAAGCTG 18502
QY 421 GGCTTCTGGGTGAGGTGGGGAATTGAGAACTTTTGTCTAGCTAAAGATTGTAATG 480
DB 18503 GGCTTCTGGGTGAGGTGGGGAATTGAGAACTTTTGTCTAGCTAAAGATTGTAATG 18562
```

```
QY 481 CACCAATAGCACTCTGTCTAGCTTAAAGATTGTAATGACCAATCAGCACTCTGTA 540
DB 18563 CACCAATAGCACTCTGTCTAGCTTAAAGATTGTAATGACCAATCAGCACTCTGTA 18622
QY 541 AAATGACCAATCAGCAGGATGTGGCGGGGCAATTAAGGAGTAAATCTGGCCACC 600
DB 18623 AAATGACCAATCAGCAGGATGTGGCGGGGCAATTAAGGAGTAAATCTGGCCACC 18682
QY 601 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660
DB 18683 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 18742
QY 661 CTCTTCAATTAATCTTGTCTGTCTGCTCATTTCTTTGTGCCACTACTTTATAGCTG 720
DB 18743 CTCTTCAATTAATCTTGTCTGTCTGCTCATTTCTTTGTGCCACTACTTTATAGCTG 18802
QY 721 TAAACCTGCTGCGGGGCTGTGGCTTCACTCTGAAGTAAACAGACCAAGCAACCACT 780
DB 18803 TAAACCTGCTGCGGGGCTGTGGCTTCACTCTGAAGTAAACAGACCAAGCAACCACT 18862
QY 781 GGAAGGAACMAAAGAACTCCCGATGTCTGCTCTTTAAGACTGTAACTCTGCGAAGC 840
DB 18863 GGAAGGAACMAAAGAACTCCCGATGTCTGCTCTTTAAGACTGTAACTCTGCGAAGC 18922
QY 841 TCTGCACTTCACTCTCTGAATGTCAGTGAGACCAACAAACCCACAGAGGAAGAACTCTG 900
DB 18923 TCTGCACTTCACTCTCTGAATGTCAGTGAGACCAACAAACCCACAGAGGAAGAACTCTG 18982
QY 901 GACACACCTGAATATCTGAAGGAACAACTCCAGACACCACTCTTTCAGACTGTAA 960
DB 18983 GACACACCTGAATATCTGAAGGAACAACTCCAGACACCACTCTTTCAGACTGTAA 19042
QY 961 CTCACCGCAAGGGTCTGTGGCTTCACTCTTGAAGTCAAGACCAACAAACCCAGGAA 1020
DB 19043 CTCACCGCAAGGGTCTGTGGCTTCACTCTTGAAGTCAAGACCAACAAACCCAGGAA 19102
QY 1021 GGAACAAATTCAGACACAGTGAAGAAATCTGATTTTGTGATCTGTGGCTTCCAGGCTTAC 1080
DB 19103 GGAACAAATTCAGACACAGTGAAGAAATCTGATTTTGTGATCTGTGGCTTCCAGGCTTAC 19162
QY 1081 TCCAGCATTTGAAGTCCATTCAGCTTGAAGCAAGAGAAAGAGTGTGGAGAGCAC 1140
DB 19163 TCCAGCATTTGAAGTCCATTCAGCTTGAAGCAAGAGAAAGAGTGTGGAGAGCAC 19222
QY 1141 ATGTGGAAATTTGTAAGACACAGGCTTGAGATGACATGAGGCAATTTGATCAAACTTA 1200
DB 19223 ATGTGGAAATTTGTAAGACACAGGCTTGAGATGACATGAGGCAATTTGATCAAACTTA 19282
QY 1201 GCTGGAAGCAGGGCCAGAAATTAATCTTAAGGAAGACAGTTTGTGACAGTATGAT 1260
DB 19283 GCTGGAAGCAGGGCCAGAAATTAATCTTAAGGAAGACAGTTTGTGACAGTATGAT 19342
QY 1261 CTTTGCATCTGAGACATGTATGATTAACAAGCAATTAATTAAGAAAATATAGCCAGGTGC 1320
DB 19343 CTTTGCATCTGAGACATGTATGATTAACAAGCAATTAATTAAGAAAATATAGCCAGGTGC 19402
QY 1321 GATGAGCTCATGCTGTAACTCCAGCACTTTGGAGGCAAGGGGTGTGATTCAGAGTTC 1380
DB 19403 GATGAGCTCATGCTGTAACTCCAGCACTTTGGAGGCAAGGGGTGTGATTCAGAGTTC 19462
QY 1381 AGGCGTTGAGACCAAGCTGGCCAAATGTGTAAACCCGCTCTTACTATAAAATACAAA 1440
DB 19463 AGGCGTTGAGACCAAGCTGGCCAAATGTGTAAACCCGCTCTTACTATAAAATACAAA 19522
QY 1441 ATTAGCTGTGTGTGTGAGCAGCATCTGTAAATCCAGTACTCAGAGGCTGAGGCGGGG 1500
DB 19523 ATTAGCTGTGTGTGTGAGCAGCATCTGTAAATCCAGTACTCAGAGGCTGAGGCGGGG 19582
QY 1501 AATCTCTGAACCTTGGAGGCAAGGTTGAGTGAAGCAAGATCAACCAAGCACTTCA 1560
DB 19583 AATCTCTGAACCTTGGAGGCAAGGTTGAGTGAAGCAAGATCAACCAAGCACTTCA 19642
QY 1561 TCCTGGGTGACAGAGCGAGACTGTCTCAAAAAAAAAAAAAAAAAAGAAAGAAAT 1620
```

Db 19643 TCCTGGGTGACAGAGCAGACTGTCTCAAAAAAAAAAAAAAAAAAGAAAGAAAT 19702  
Qy 1621 ATATCAAGATATTTGACAGTAACTTTATTCACACTTATGACACAGCAATACA 1680  
Db 19703 ATATCAAGATATTTGACAGTAACTTTATTCACACTTATGACACAGCAATACA 19762  
Qy 1681 CTAGTGTATTTACATGATTAACCTTATTCATTAACAATAGCCCTATGAGTACAGTGC 1740  
Db 19763 CTAGTGTATTTACATGATTAACCTTATTCATTAACAATAGCCCTATGAGTACAGTGC 19822  
Qy 1741 TGTATATATCTCACTTATATAGTAAGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 1800  
Db 19823 TGTATATATCTCACTTATATAGTAAGAAACTGAAGTACAGAAAGGTCAAGTAGAGAAA 19882  
Qy 1801 TGGCCATGCTTGATCTGCTGTTTGAAGCACTGTAACAGAAATCGGTGAGAAAT 1860  
Db 19883 TGGCCATGCTTGATCTGCTGTTTGAAGCACTGTAACAGAAATCGGTGAGAAAT 19942  
Qy 1861 GCTCTAACAAGATGTAGTACAGGGGTTGGAGGTAAGTGTGAGTGGGAGTTGGGG 1920  
Db 19943 GCTCTAACAAGATGTAGTACAGGGGTTGGAGGTAAGTGTGAGTGGGAGTTGGGG 20002  
Qy 1921 ATGGAAGATGATGAGAAACAGCTTGACAGAGAGCTGACACTGGCACTCTGGGGA 1980  
Db 20003 ATGGAAGATGATGAGAAACAGCTTGACAGAGAGCTGACACTGGCACTCTGGGGA 20062  
Qy 1981 CCTGAAGGTTAGAGGACT 2001  
Db 20063 CCTGAAGGTTAGAGGACT 20083

RESULT 2  
AF222855 80622 bp DNA linear PRI 17-JUL-2000  
LOCUS Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.  
DEFINITION AF222855  
ACCESSION AF222855.1 GI:9246845  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 80622)  
AUTHORS Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,  
Irvine,D.V., Safefy,R. and Choo,K.H.  
TITLE The 10q25 neocentromere and its inactive progenitor have identical  
primary nucleotide sequence: further evidence for epigenetic  
modification  
JOURNAL Genome Res. 10 (6), 832-838 (2000)  
PUBMED 10854414  
REFERENCE 2 (bases 1 to 80622)  
AUTHORS Barry,A.E.  
TITLE Direct Submission  
JOURNAL Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch  
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,  
Parkville,, Melbourne, Victoria 3052, Australia  
REMARK Genomic sequence from human 10q25.2, clone1b=HC  
FEATURES  
Location/Qualifiers  
1. 80622  
source  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10q25.2"  
/clone="HC"  
repeat\_region complement(3, 176)  
/rpt\_family="Alu"  
repeat\_region complement(979, 3564)  
/rpt\_family="Tiggeci"  
2967, 3094  
repeat\_region /rpt\_family="Alu"  
/rpt\_family="Alu"

repeat\_region 3664, 3934  
/rpt\_family="Alu"  
repeat\_region complement(4920, 5181)  
/rpt\_family="Alu"  
repeat\_region complement(6287, 6579)  
/rpt\_family="Alu"  
repeat\_region 6625, 6966  
/rpt\_family="MLT2B2"  
repeat\_region complement(9215, 9490)  
/rpt\_family="Alu"  
repeat\_region complement(11280, 11565)  
/rpt\_family="Alu"  
repeat\_region complement(12428, 12675)  
/rpt\_family="Alu"  
repeat\_region complement(14304, 14583)  
/rpt\_family="Alu"  
repeat\_region 17503, 17794  
/rpt\_family="Alu"  
repeat\_region complement(20250, 20533)  
/rpt\_family="Alu"  
repeat\_region complement(23031, 23323)  
/rpt\_family="Alu"  
repeat\_region complement(24301, 24479)  
/rpt\_family="MIR"  
repeat\_region 24951, 25250  
/rpt\_family="Alu"  
repeat\_region complement(25987, 26342)  
/rpt\_family="TBR1"  
repeat\_region 27097, 27335  
/rpt\_family="MER2"  
repeat\_region 27538, 27814  
/rpt\_family="Alu"  
repeat\_region 27891, 28161  
/rpt\_family="Alu"  
repeat\_region 29063, 29334  
/rpt\_family="Alu"  
repeat\_region complement(31025, 31093)  
/rpt\_family="MLT1"  
repeat\_region complement(31817, 32352)  
/rpt\_family="MER1"  
repeat\_region complement(32489, 32607)  
/rpt\_family="MLT1"  
repeat\_region 33035, 33316  
/rpt\_family="Alu"  
repeat\_region 33469, 33760  
/rpt\_family="Alu"  
repeat\_region complement(34527, 34590)  
/rpt\_family="MIR"  
repeat\_region 34685, 34995  
/rpt\_family="Alu"  
repeat\_region complement(36349, 36437)  
/rpt\_family="MIR"  
repeat\_region complement(38796, 39079)  
/rpt\_family="Alu"  
repeat\_region 39441, 39693  
/rpt\_family="Alu"  
repeat\_region complement(43176, 43237)  
/rpt\_family="MIR"  
repeat\_region complement(45203, 45478)  
/rpt\_family="Alu"  
repeat\_region 45892, 46025  
/rpt\_family="MIR"  
repeat\_region complement(46618, 46851)  
/rpt\_family="Alu"  
repeat\_region 48623, 48939  
/rpt\_family="Alu"  
repeat\_region complement(49048, 50698)  
/rpt\_family="L1"  
repeat\_region complement(50710, 50972)  
/rpt\_family="Alu"  
repeat\_region complement(51031, 51192)  
/rpt\_family="L1"  
repeat\_region complement(51450, 51739)



```
repeat_region /rpt_family="Alu" complement(53270, .53562)
repeat_region /rpt_family="Alu" 54261, .54808
repeat_region /rpt_family="LTR12" 59460, .60054
repeat_region /rpt_family="LTR12" 60345, .60633
repeat_region /rpt_family="Alu" complement(60669, .60801)
repeat_region /rpt_family="MIR" complement(61684, .61852)
repeat_region /rpt_family="MERS" 63466, .63753
repeat_region /rpt_family="Alu" 64063, .64228
repeat_region /rpt_family="MERS" 64446, .64521
repeat_region /rpt_family="MIR" 65794, .65883
repeat_region /rpt_family="MERS" 66772, .66951
repeat_region /rpt_family="MER20" complement(67895, .68416)
repeat_region /rpt_family="MER41" complement(73192, .73470)
repeat_region /rpt_family="Alu" 74086, .74367
repeat_region /rpt_family="Alu" 75916, .76183
repeat_region /rpt_family="Alu" complement(77771, .78059)
repeat_region /rpt_family="Alu"
```

## ORIGIN

Query Match 100.0%; Score 2001; DB 8; Length 80622;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gap 0;

```
QY 1 AAGGCTTGAATCGACATGCTTCAAACTTTATACCACTTGGAGTTGG 60
DB 59027 AAGGCTTGAATCGACATGCTTCAAACTTTATACCACTTGGAGTTGG 59086
QY 61 GACATGGCTTCTCCCTTTCTAGGTCTGTGACAGCCATCTTGTAATGCGATTTG 120
DB 59087 GACATGGCTTCTCCCTTTCTAGGTCTGTGACAGCCATCTTGTAATGCGATTTG 59146
QY 121 GCCCTGATTTTAACTCTTGGTCAAAATTTGTTTCTCTAGATCGAGGCAATCAAGT 180
DB 59147 GCCCTGATTTTAACTCTTGGTCAAAATTTGTTTCTCTAGATCGAGGCAATCAAGT 59206
QY 181 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCAACTAACACTTGTCTGAGAGC 240
DB 59207 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCAACTAACACTTGTCTGAGAGC 59266
QY 241 CTTGAGACCGACCGCTGACCTTTCATGAGCTTAAAGCTCCCTCTGAGAGCACTAC 300
DB 59267 CTTGAGACCGACCGCTGACCTTTCATGAGCTTAAAGCTCCCTCTGAGAGCACTAC 59326
QY 301 CACTGACGAGGCGCTTCTTCACCCTATCCAGACGAGAGTAGCTACAGCGGTCAATCGCA 360
DB 59327 CACTGACGAGGCGCTTCTTCACCCTATCCAGACGAGAGTAGCTACAGCGGTCAATCGCA 59386
QY 361 AATCCCAACAGACGCTGGGTGTCTCTGTTTGAAGGGGGAATTGAAGGTGAAGCCAGCTG 420
DB 59387 AATCCCAACAGACGCTGGGTGTCTCTGTTTGAAGGGGGAATTGAAGGTGAAGCCAGCTG 59446
QY 421 GGCTTCTGGGTCAAGTGGGGAATTGAGAACCTTTGTCTTAGCTAAAGATTGTAATG 480
DB 59447 GGCTTCTGGGTCAAGTGGGGAATTGAGAACCTTTGTCTTAGCTAAAGATTGTAATG 59506
QY 481 CACCAATAGCACTCTGTCTAGCTAAAGATTGTAATGACCAATCAGACACTCTGTA 540
DB CACCAATAGCACTCTGTCTAGCTAAAGATTGTAATGACCAATCAGACACTCTGTA 540
```

```
DB 59507 CACCAATAGCACTCTGTCTAGCTAAAGATTGTAATGACCAATCAGACACTCTGTA 59566
QY 541 AATAGACCAATCAGACGATGTGGCGGGTCAATATAGGAGTAAAACTGGCCACCC 600
DB 59567 AATAGACCAATCAGACGATGTGGCGGGTCAATATAGGAGTAAAACTGGCCACCC 59626
QY 601 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCCACTGTGGAAGCTTGTCTTTTG 660
DB 59627 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCCACTGTGGAAGCTTGTCTTTTG 59686
QY 661 CTCTTCACAATTAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 59687 CTCTTCACAATTAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 59746
QY 721 TAACACTCACTGAGAGGTCTGTGCTTCAATCTGAACTCAACAGACCAAGCCACT 780
DB 59747 TAACACTCACTGAGAGGTCTGTGCTTCAATCTGAACTCAACAGACCAAGCCACT 59806
QY 781 GGAAGGACAAAGAACTCCGATGTGTGCTCTTTAAGAGCTGTAACTCACTGCGAAGC 840
DB 59807 GGAAGGACAAAGAACTCCGATGTGTGCTCTTTAAGAGCTGTAACTCACTGCGAAGC 59866
QY 841 TCTGACGCTTCACTCCTGAAAGTCAAGAGACCAAAACCCACAGAAAGAAACTCTG 900
DB 59867 TCTGACGCTTCACTCCTGAAAGTCAAGAGACCAAAACCCACAGAAAGAAACTCTG 59926
QY 901 GACACACTGTAATATCTGAAGAACAACTCCAGACACACTTTTCAGAGCTGTAA 960
DB 59927 GACACACTGTAATATCTGAAGAACAACTCCAGACACACTTTTCAGAGCTGTAA 59986
QY 961 CTCACCGCAAGGCTGTGGCTTCAATCTTTGAAGTACGAACCAAGAACCAACCGGAA 1020
DB 59987 CTCACCGCAAGGCTGTGGCTTCAATCTTTGAAGTACGAACCAAGAACCAACCGGAA 60046
QY 1021 GGAACCAATTCAGACAGTGGGAATCTGTAATTTTATGATCTGTGGCTTCCAGGTTAC 1080
DB 60047 GGAACCAATTCAGACAGTGGGAATCTGTAATTTTATGATCTGTGGCTTCCAGGTTAC 60106
QY 1081 TCCAGTCAATTAAGTCTCCATTCAGCTTAAAGAAACAGAGATGTTTGAAGACAC 1140
DB 60107 TCCAGTCAATTAAGTCTCCATTCAGCTTAAAGAAACAGAGATGTTTGAAGACAC 60166
QY 1141 ATGTGGAAATGTTATGAGACAGGCTTGAATCAATAGGCAATTTCTGATCAACCTA 1200
DB 60167 ATGTGGAAATGTTATGAGACAGGCTTGAATCAATAGGCAATTTCTGATCAACCTA 60226
QY 1201 GCTGGAAGCAGGGCCAGAAATTAATCTTAAGGAAGACGTTTGTAGACAGTAGT 1260
DB 60227 GCTGGAAGCAGGGCCAGAAATTAATCTTAAGGAAGACGTTTGTAGACAGTAGT 60286
QY 1261 CTTTGCACTGAGACATGTAGATTATCAAGCAATTAATTAAGAAAAATATAGCCAGTGC 1320
DB 60287 CTTTGCACTGAGACATGTAGATTATCAAGCAATTAATTAAGAAAAATATAGCCAGTGC 60346
QY 1321 GATGGCTCATGCTGTATATCCACACTTTTGGAGGCCAAGGGGTGTGATCACGAGTC 1380
DB 60347 GATGGCTCATGCTGTATATCCACACTTTTGGAGGCCAAGGGGTGTGATCACGAGTC 60406
QY 1381 AGCGCTTGAGACCAAGCTTGCCCAATGTGTGAACCCCGCTCTACTTAAAAATACAAA 1440
DB 60407 AGCGCTTGAGACCAAGCTTGCCCAATGTGTGAACCCCGCTCTACTTAAAAATACAAA 60466
QY 1441 ATTGACCTGTGTGTGTGACGACTGTATATCCAGTAATCTCAGAGAGCTGAGCAGGG 1500
DB 60467 ATTGACCTGTGTGTGTGACGACTGTATATCCAGTAATCTCAGAGAGCTGAGCAGGG 60526
QY 1501 AATCTTTGAACCTTGGAGGAGAGGTTGACAGTGAAGCCAGATCAACACAGCACTTCA 1560
DB 60527 AATCTTTGAACCTTGGAGGAGAGGTTGACAGTGAAGCCAGATCAACACAGCACTTCA 60586
QY 1561 TCTGTGGTGAAGAGCAACTCTGTCTCAAAAAAATTTTAAAAAAGAAAGAAAT 1620
DB 60587 TCTGTGGTGAAGAGCAACTCTGTCTCAAAAAAATTTTAAAAAAGAAAGAAAT 60646
```

QY 1621 ATATCAAGAAATTTGACAGTAACTTTATTCACACTATGACACGAGCAATAC 1680  
|  
|  
|  
Db 60647 ATATCAAGAAATTTGACAGTAACTTTATTCACACTATGACACGAGCAATAC 60706  
|  
|  
|  
QY 1681 CTAGTGTGTTTACATGATTAACCTATTATCTTAAACAATAGCCCTATGAGTGC 1740  
|  
|  
|  
Db 60707 CTAGTGTGTTTACATGATTAACCTATTATCTTAAACAATAGCCCTATGAGTGC 60766  
|  
|  
|  
QY 1741 TGTATTATCTCCACTTTTATGATTAAGAACTGAAGTACAGAAAGTCAAGTGA 1800  
|  
|  
|  
Db 60767 TGTATTATCTCCACTTTTATGATTAAGAACTGAAGTACAGAAAGTCAAGTGA 60826  
|  
|  
|  
QY 1801 TGGCCATGCTTGCACTCTGAGTTTGAAGCACTGTTACAGAACTGGTGTGAGAAAT 1860  
|  
|  
|  
Db 60827 TGGCCATGCTTGCACTCTGAGTTTGAAGCACTGTTACAGAACTGGTGTGAGAAAT 60886  
|  
|  
|  
QY 1861 GCTTACACAGATGTAGTCAAGGAGTGGAGGTAAGTCTGAGTTGGGAGTTGGG 1920  
|  
|  
|  
Db 60887 GCTTACACAGATGTAGTCAAGGAGTGGAGGTAAGTCTGAGTTGGGAGTTGGG 60946  
|  
|  
|  
QY 1921 ATGGAAGATGATGTAAGAAACAGCTTGACAGAGACCTGCACTTGGCACTGTGGGA 1980  
|  
|  
|  
Db 60947 ATGGAAGATGATGTAAGAAACAGCTTGACAGAGACCTGCACTTGGCACTGTGGGA 61006  
|  
|  
|  
QY 1981 CCTGAAAGGTTTGAAGGACT 2001  
|  
|  
|  
Db 61007 CCTGAAAGGTTTGAAGGACT 61027  
|  
|  
|

RESULT 3  
AX033939 39238 bp DNA linear PAT 21-SEP-2000

LOCUS AX033939 39238 bp DNA linear PAT 21-SEP-2000  
DEFINITION Sequence 31 from Patent WO9511790.  
ACCESSION AX033939  
VERSION AX033939.1 GI:10280508  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
unclassified sequences.

REFERENCE  
1 Cancelli, M.R., Choo, K.H. and Du, S.D.  
A novel nucleic acid molecule  
Patent: WO 9851790-A 31 19-NOV-1998;  
JOURNAL CANCELLI MICHAEL ROBERT (AU); CHOO KONG HONG ANDY (AU); SART  
DESIREE DU (AU); AMRAD OPERATIONS PTY LTD (AU)  
Location/Qualifiers

FEATURES  
source  
1..39238  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 99.9%; Score 1999.4; DB 6; Length 39238;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAGCTTGAATTCAGCAATGCTTTCAACCTCTTATACCACTCTGAGTTGGC 60  
|  
|  
|  
Db 17759 AAAAGCTTGAATTCAGCAATGCTTTCAACCTCTTATACCACTCTGAGTTGGC 17818  
|  
|  
|  
QY 61 GACATGGCTTCCCTCTTCTGAGTCTGTGACAGCACTTGTCTATATGTCGCAATTTGG 120  
|  
|  
|  
Db 17819 GACATGGCTTCCCTCTTCTGAGTCTGTGACAGCACTTGTCTATATGTCGCAATTTGG 17878  
|  
|  
|  
QY 121 GCCCTGATTTTAACTCTGATCAATTTGTTCTCTAGAGTGAAGCCATCAAGCT 180  
|  
|  
|  
Db 17879 GCCCTGATTTTAACTCTGATCAATTTGTTCTCTAGAGTGAAGCCATCAAGCT 17938  
|  
|  
|  
QY 181 ACAGATGATCTTACAAATGTAAACCCCAATGAGCTCACTTCTGCTGAGGACC 240  
|  
|  
|  
Db 17939 ACAGATGATCTTACAAATGTAAACCCCAATGAGCTCACTTCTGCTGAGGACC 17998  
|  
|  
|

QY 241 CCTGACCAAGCCGCTGACCCTTTCAATGAGCCTAAAGAGCTCCCTCTGAGAGCACTAC 300  
|  
|  
|  
Db 17999 CCTGACCAAGCCGCTGACCCTTTCAATGAGCCTAAAGAGCTCCCTCTGAGAGCACTAC 18058  
|  
|  
|  
QY 301 CACTGAGGAGCCCTTTCTCAACCCCTATCCAGAGGAATGATCAAGCGGTATGCGCA 360  
|  
|  
|  
Db 18059 CACTGAGGAGCCCTTTCTCAACCCCTATCCAGAGGAATGATCAAGCGGTATGCGCA 18118  
|  
|  
|  
QY 361 AATCCCAAGAGAGCTGGGGGTGCTCTGTTTGAAGGGGGATTTGAAGGTGAAGCCAGCT 420  
|  
|  
|  
Db 18119 AATCCCAAGAGAGCTGGGGGTGCTCTGTTTGAAGGGGGATTTGAAGGTGAAGCCAGCT 18178  
|  
|  
|  
QY 421 GCTTCTGGGTCAAGTGGGAGCACTTGAGAACTTTTGTCTAGCTTAAAGATTTAAATG 480  
|  
|  
|  
Db 18179 GCTTCTGGGTCAAGTGGGAGCACTTGAGAACTTTTGTCTAGCTTAAAGATTTAAATG 18238  
|  
|  
|  
QY 481 CACCAATCAGCACTGTGTCTAGCTTAAAGATTTGTAATGCAACAATGAGCACTGTGTA 540  
|  
|  
|  
Db 18239 CACCAATCAGCACTGTGTCTAGCTTAAAGATTTGTAATGCAACAATGAGCACTGTGTA 18298  
|  
|  
|  
QY 541 AATGAGCAATTCAGAGATGTGGGCGGGGTCAAATTAAGGAGTAAAACTGGCCACCC 600  
|  
|  
|  
Db 18299 AATGAGCAATTCAGAGATGTGGGCGGGGTCAAATTAAGGAGTAAAACTGGCCACCC 18358  
|  
|  
|  
QY 601 GAGCCAGAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660  
|  
|  
|  
Db 18359 GAGCCAGAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 18418  
|  
|  
|  
QY 661 CTCTTCAATTAATCTTGTCTGCTGCTCACTTCTGTGTGTCACACTTACTTATAGACTG 720  
|  
|  
|  
Db 18419 CTCTTCAATTAATCTTGTCTGCTGCTCACTTCTGTGTGTCACACTTACTTATAGACTG 18478  
|  
|  
|  
QY 721 TAACTCTCACTGCGAGGCTGTGTGCTTATCTCTGAAGTCAACAGACCAAGCCACT 780  
|  
|  
|  
Db 18479 TAACTCTCACTGCGAGGCTGTGTGCTTATCTCTGAAGTCAACAGACCAAGCCACT 18538  
|  
|  
|  
QY 781 GGAAGGAACAAAGAACTCCCGATGTGCTGCTTTAAGAGCTGAACACTCATGCGGAAC 840  
|  
|  
|  
Db 18539 GGAAGGAACAAAGAACTCCCGATGTGCTGCTTTAAGAGCTGAACACTCATGCGGAAC 18598  
|  
|  
|  
QY 841 TCTGAGCTTCACTCTGTAAGTCAGTGAGACCAACCCACAGAGGAAGAACTCTG 900  
|  
|  
|  
Db 18599 TCTGAGCTTCACTCTGTAAGTCAGTGAGACCAACCCACAGAGGAAGAACTCTG 18658  
|  
|  
|  
QY 901 GACACCTGAAATCTGTAAGGAAGAACTCCAGACACCACTTTTCAAGCTGTAA 960  
|  
|  
|  
Db 18659 GACACCTGAAATCTGTAAGGAAGAACTCCAGACACCACTTTTCAAGCTGTAA 18718  
|  
|  
|  
QY 961 CTCACCGCAAGGGTCTGAGCTTCACTTCTTAAGTCAAGAAACCAAGAACCCACCGGA 1020  
|  
|  
|  
Db 18719 CTCACCGCAAGGGTCTGAGCTTCACTTCTTAAGTCAAGAAACCAAGAACCCACCGGA 18778  
|  
|  
|  
QY 1021 GGAACAAATTCAGACAGTGAAGAAATCTGATTTTGTGAGTCTGAGGTTAC 1080  
|  
|  
|  
Db 18779 GGAACAAATTCAGACAGTGAAGAAATCTGATTTTGTGAGTCTGAGGTTAC 18838  
|  
|  
|  
QY 1081 TCCAGTCAATGAAGTCTCAATTCAGCTTAAAGAAACAGAGATGTTGAGAGACAC 1140  
|  
|  
|  
Db 18839 TCCAGTCAATGAAGTCTCAATTCAGCTTAAAGAAACAGAGATGTTGAGAGAGAC 18898  
|  
|  
|  
QY 1141 ATGTGGAAATCTTATGACCAAGGCTTGAAGACCATAGGGCACTTCTGATCAAACTTA 1200  
|  
|  
|  
Db 18899 ATGTGGAAATCTTATGACCAAGGCTTGAAGACCATAGGGCACTTCTGATCAAACTTA 18958  
|  
|  
|  
QY 1201 GCTGAGACAGGGCCAGAGAAATATATCTAAGAGAGACAGTTTGTGAGACGTAGT 1260  
|  
|  
|  
Db 18959 GCTGAGACAGGGCCAGAGAAATATATCTAAGAGAGACAGTTTGTGAGACGTAGT 19018  
|  
|  
|  
QY 1261 CTTTGCATCTGAGACATGTAGTATCAAGCAATTAATTAGAAAAATATAGCCAGGTGC 1320  
|  
|  
|  
Db 19019 CTTTGCATCTGAGACATGTAGTATCAAGCAATTAATTAGAAAAATATAGCCAGGTGC 19078  
|  
|  
|  
QY 1321 GATGCTCATGCTCTTAATCCAGACACTTTGGAGAGCCAGGGGTGTGATCAGAGGT 1380  
|  
|  
|

|||||  
Db 19079 GATGGCTATGCTCTTAATCCAGCATTTGGAGGCGAAGGGGTGGATCACAAGTTC 19138  
|||  
Qy 1381 AGGCGTTGAGACGACGCTGGCCAACTGTGTAAACCCCGTCTCTACTTAAAAATACAAA 1440  
|||  
Db 19139 AGGCGTTGAGACGACGCTGGCCAACTGTGTAAACCCCGTCTCTACTTAAAAATACAAA 19198  
|||  
Qy 1441 ATTACCGTGTGTGTGGACGATCTGTAAATCCCAATCTAGAGAGGCTGAGGAGGG 1500  
|||  
Db 19199 ATTACCGTGTGTGTGGACGATCTGTAAATCCCAATCTAGAGAGGCTGAGGAGGG 19258  
|||  
Qy 1501 AATCTCTTGAATTTGGGAGGCGAGGTTGACGTGAGCCAAAGTTCACACGACACTCCA 1560  
|||  
Db 19259 AATCTCTTGAATTTGGGAGGCGAGGTTGACGTGAGCCAAAGTTCACACGACACTCCA 19318  
|||  
Qy 1561 TCTCGGTGACAGAGCGAGACTCTGTCTCTCAAAAAAAAAAAAAAAAAAGAAAT 1620  
|||  
Db 19319 TCTCGGTGACAGAGCGAGACTCTGTCTCTCAAAAAAAAAAAAAAAAAAGAAAT 19378  
|||  
Qy 1621 AATATCAGAAATTTGACAGGTAACATTTATTCACACTTACTATGCAAGCAATACA 1680  
|||  
Db 19379 AATATCAGAAATTTGACAGGTAACATTTATTCACACTTACTATGCAAGCAATACA 19438  
|||  
Qy 1681 CTTAAGGTTTTCATGAGATTAATCTCAATTTATCTTAAACAATAGCCCTATGAGTCACTGC 1740  
|||  
Db 19439 CTTAAGGTTTTCATGAGATTAATCTCAATTTATCTTAAACAATAGCCCTATGAGTCACTGC 19498  
|||  
Qy 1741 TGTATTAATCTCCACTTATAGATTAAGAAACTGAAAGTACAGAAAGTCAAGTAGAGAA 1800  
|||  
Db 19499 TGTATTAATCTCCACTTATAGATTAAGAAACTGAAAGTACAGAAAGTCAAGTAGAGAA 19558  
|||  
Qy 1801 TGGCCATGCTTGATCTCAGTTTGTGAAGCACTGTACAGAAATCTGGTGTGAGAAAT 1860  
|||  
Db 19559 TGGCCATGCTTGATCTCAGTTTGTGAAGCACTGTACAGAAATCTGGTGTGAGAAAT 19618  
|||  
Qy 1861 GCTCTTAAACAAGTGTGAGGGGTTGGAGGTAAGTCTGAGTTGGCAGTTGGG 1920  
|||  
Db 19619 GCTCTTAAACAAGTGTGAGGGGTTGGAGGTAAGTCTGAGTTGGCAGTTGGG 19678  
|||  
Qy 1921 ATGGAAGATGATGTAAGAAACAGTTGACAGAGAACTGACACTTGGCAACTCTGTGGGA 1980  
|||  
Db 19679 ATGGAAGATGATGTAAGAAACAGTTGACAGAGAACTGACACTTGGCAACTCTGTGGGA 19738  
|||  
Qy 1981 CTTTGAAGGTTAGGGGACT 2001  
|||  
Db 19739 CTTTGAAGGTTAGGGGACT 19759  
|||

RESULT 4  
AF222856 69058 bp DNA linear PRI 17-JUN-2000  
LOCUS Homo sapiens clone Pnc chromosome 10 map 10q25.2 genomic sequence.  
DEFINITION AF222856  
ACCESSION AF222856.1 GI:9246846  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo  
1 (bases 1 to 69058)  
Barry,A.E., Bateman,M., Howman,E.V., Cancelli,M.R., Tainton,K.M.,  
Irvine,D.V., Saffery,R. and Cho,K.H.  
The 10q25 neocentromere and its inactive progenitor have identical  
primary nucleotide sequence: further evidence for epigenetic  
modification  
Genome Res. 10 (6), 832-838 (2000)  
JOURNAL PUBMED  
10854414  
2 (bases 1 to 69058)  
REFERENCE AUTHORS Barry,A.E.  
TITLE Direct Submission  
JOURNAL Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch  
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,

REMARK Parkville,, Melbourne, Victoria 3052, Australia  
FEATURES Human genomic sequence from 10q25.2, cloneLib=Pnc  
SOURCE Location/Qualifiers  
1. 69058  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10q25.2"  
/clone="Pnc"  
complement(598..3183)  
/rpt\_family="T199er1"  
2587..2715  
/rpt\_family="Alu"  
3283..3553  
/rpt\_family="Alu"  
complement(4548..4809)  
/rpt\_family="Alu"  
complement(5919..6211)  
/rpt\_family="Alu"  
6257..6599  
/rpt\_family="MLT2B2"  
complement(8848..9123)  
/rpt\_family="Alu"  
complement(10913..11198)  
/rpt\_family="Alu"  
complement(12062..12313)  
/rpt\_family="Alu"  
complement(13957..14236)  
/rpt\_family="Alu"  
17133..17424  
/rpt\_family="Alu"  
complement(19882..20197)  
/rpt\_family="Alu"  
complement(22654..22956)  
/rpt\_family="Alu"  
complement(23935..24113)  
/rpt\_family="MIR"  
24585..24884  
/rpt\_family="Alu"  
complement(25618..25973)  
/rpt\_family="THER1"  
26732..26970  
/rpt\_family="MER2"  
27173..27449  
/rpt\_family="Alu"  
27526..27796  
/rpt\_family="Alu"  
28702..28978  
/rpt\_family="Alu"  
complement(30668..30736)  
/rpt\_family="MLT1"  
complement(31460..31993)  
/rpt\_family="MER1"  
complement(32130..32248)  
/rpt\_family="MLT1"  
32685..32961  
/rpt\_family="Alu"  
33113..33403  
/rpt\_family="Alu"  
complement(34169..34233)  
/rpt\_family="MIR"  
34332..34639  
/rpt\_family="Alu"  
complement(35993..36081)  
/rpt\_family="MIR"  
complement(38438..38721)  
/rpt\_family="Alu"  
39081..39333  
/rpt\_family="Alu"  
complement(42817..42877)  
/rpt\_family="MIR"  
complement(44845..45119)

/rpt\_family="Alu"  
repeat\_region 45532..45665  
/rpt\_family="MIR"  
repeat\_region complement(46257)..46490  
/rpt\_family="Alu"  
repeat\_region 48262..48578  
/rpt\_family="Alu"  
repeat\_region complement(48687)..50337  
/rpt\_family="L1"  
repeat\_region complement(50349)..50611  
/rpt\_family="Alu"  
repeat\_region complement(50670)..50831  
/rpt\_family="L1"  
repeat\_region complement(51085)..51374  
/rpt\_family="Alu"  
repeat\_region complement(52905)..53197  
/rpt\_family="Alu"  
repeat\_region 53896..54467  
/rpt\_family="LTR12"  
repeat\_region 59129..59723  
/rpt\_family="LTR12"  
repeat\_region 60018..60302  
/rpt\_family="Alu"  
repeat\_region complement(60338)..60470  
/rpt\_family="MIR"  
repeat\_region complement(61353)..61521  
/rpt\_family="MERS"  
repeat\_region 63135..63416  
/rpt\_family="Alu"  
repeat\_region 63724..63889  
/rpt\_family="MERS"  
repeat\_region 64107..64182  
/rpt\_family="MIR"  
repeat\_region 65435..65548  
/rpt\_family="MERS"  
repeat\_region 66439..66618  
/rpt\_family="MER20"  
repeat\_region complement(67562)..68084  
/rpt\_family="MER41"

## ORIGIN

Query Match 99.9%; Score 1999.4; DB 8; Length 69058;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAGCTTGTGAATCAGACAAATGCTTTCAAACTTTATACCACCTTGAAGTTGGG 60  
DB 58696 AAAGCTTGTGAATCAGACAAATGCTTTCAAACTTTATACCACCTTGAAGTTGGG 58755  
QY 61 GACATGGCTTCCCTCTTCTAGTGTCTGTGACAGGCATCTGTAATAGTGGCATTTGG 120  
DB 58756 GACATGGCTTCCCTCTTCTAGTGTCTGTGACAGGCATCTGTAATAGTGGCATTTGG 58815  
QY 121 GCCCTGATTTTAACTCTGTGTAATTTGTTCTCTAGATCGAGCCATCAAGCT 180  
DB 58816 GCCCTGATTTTAACTCTGTGTAATTTGTTCTCTAGATCGAGCCATCAAGCT 58875  
QY 181 ACAGATGATCTTCAAAATGTAAACCCCAAATGAGCTCACTAACACTTCTGTGAGAAC 240  
DB 58876 ACAGATGATCTTCAAAATGTAAACCCCAAATGAGCTCACTAACACTTCTGTGAGAAC 58935  
QY 241 CCTGACCGACCGCGCTGGCCCTTTCAATGGCCCTTAAGAGCTCCCTCTGGAAGACATAC 300  
DB 58936 CCTGACCGACCGCGCTGGCCCTTTCAATGGCCCTTAAGAGCTCCCTCTGGAAGACATAC 58995  
QY 301 CACTGACGGGCCCTTCTTACCCCTATCCAGCAGGAAATAGCTACAGGGTCAATCGCA 360  
DB 58996 CACTGACGGGCCCTTCTTACCCCTATCCAGCAGGAAATAGCTACAGGGTCAATCGCA 59055  
QY 361 AATCCCAAGACGCTGGGGTGTCTGTTTGAAGGGGGGATTGAGAGGTGAAGCCAGCTG 420  
DB 59056 AATCCCAAGACGCTGGGGTGTCTGTTTGAAGGGGGGATTGAGAGGTGAAGCCAGCTG 59115

QY 421 GGCTTGGGTCAGGTGGGACCTTGGAGAACTTTTGTGTCTAGCTAAAGATTGTAAATG 480  
DB 59116 GGCTTGGGTCAGGTGGGACCTTGGAGAACTTTTGTGTCTAGCTAAAGATTGTAAATG 59175  
QY 481 CACCATTCAGCACTCTGTGTCTAGCTAAAGATTGTAAATGACCAATCAGCACTCTGTG 540  
DB 59176 CACCATTCAGCACTCTGTGTCTAGCTAAAGATTGTAAATGACCAATCAGCACTCTGTG 59235  
QY 541 AATGAGCAATCAGCAGGATGAGCGGGGGTCAATTAAGGAGTAAATCTGGCACC 600  
DB 59236 AATGAGCAATCAGCAGGATGAGCGGGGGTCAATTAAGGAGTAAATCTGGCACC 59295  
QY 601 GAGCCAGCAGTGGCAACCACTCGGCTCCCTTCCAACCTGTGGAAGCTTTGTTCTTTTG 660  
DB 59296 GAGCCAGCAGTGGCAACCACTCGGCTCCCTTCCAACCTGTGGAAGCTTTGTTCTTTTG 59355  
QY 661 CTCTTCAATTAATTTTGTGTGTCTGCTCATTTCTTTGTTCACACTACTTTATGAGCTG 720  
DB 59356 CTCTTCAATTAATTTTGTGTGTCTGCTCATTTCTTTGTTCACACTACTTTATGAGCTG 59415  
QY 721 TAAACTCACTGGGAGGGTCTGTGCTTCAATTCCTGAAGTCAACAGACCAAGCACT 780  
DB 59416 TAAACTCACTGGGAGGGTCTGTGCTTCAATTCCTGAAGTCAACAGACCAAGCACT 59475  
QY 781 GGAAGGAACAAAGAACTCCCGATGTGTCTCTTTAAGACTGTAACTCACTGCGAAGC 840  
DB 59476 GGAAGGAACAAAGAACTCCCGATGTGTCTCTTTAAGACTGTAACTCACTGCGAAGC 59535  
QY 841 TCTGAGCTTCACTCTCTGAAGTCACTGAGACCAACCAACCAAGGAAGAACTCTG 900  
DB 59536 TCTGAGCTTCACTCTCTGAAGTCACTGAGACCAACCAACCAAGGAAGAACTCTG 59595  
QY 901 GACACACCTGAATATTTGAAGGAACAACTCCAGACACACATCTTTCAAGACTGTAAAC 960  
DB 59596 GACACACCTGAATATTTGAAGGAACAACTCCAGACACACATCTTTCAAGACTGTAAAC 59655  
QY 961 CTCACCGCAAGGGTCTGTGGCTTCACTTTGAAGTCAAGGAACCAAGCAACCGGAA 1020  
DB 59656 CTCACCGCAAGGGTCTGTGGCTTCACTTTGAAGTCAAGGAACCAAGCAACCGGAA 59715  
QY 1021 GGAACAAATTCAGACACAGTGAAGAAATCTGTATTTTGTGATCTGTGGCTTCAAGGTTAC 1080  
DB 59716 GGAACAAATTCAGACACAGTGAAGAAATCTGTATTTTGTGATCTGTGGCTTCAAGGTTAC 59775  
QY 1081 TCCAGTCATTTGAAGTCCATTTGACCTTTAAGAAACAGAAATGTTTGAAGACAC 1140  
DB 59776 TCCAGTCATTTGAAGTCCATTTGACCTTTAAGAAACAGAAATGTTTGAAGACAC 59835  
QY 1141 ATGTGGAAATTTGTATGAACAGAGCTTGAGTACATAGGGCATTTCTGATCAACCTA 1200  
DB 59836 ATGTGGAAATTTGTATGAACAGAGCTTGAGTACATAGGGCATTTCTGATCAACCTA 59895  
QY 1201 GCTGGAAGCAGGGCCAGGAATATATCTTAAGAAACAGTTTGTATGACAGTATAGT 1260  
DB 59896 GCTGGAAGCAGGGCCAGGAATATATCTTAAGAAACAGTTTGTATGACAGTATAGT 59955  
QY 1261 CTTTGATCTTGAACATGTATATCAAGCAATTAATTAAGAAAAATATAGCAGGTG 1320  
DB 59956 CTTTGATCTTGAACATGTATATCAAGCAATTAATTAAGAAAAATATAGCAGGTG 60015  
QY 1321 GATGGCTATGCTGTAAATCCAGACCTTGGGAAGGCCAAAGGGGTGTGATCAAGAGTC 1380  
DB 60016 GATGGCTATGCTGTAAATCCAGACCTTGGGAAGGCCAAAGGGGTGTGATCAAGAGTC 60075  
QY 1381 AGGCGTTGAGACAGAGCTGTGGCAACATGATGAACCCCGTCTCTAATAAATAACAAA 1440  
DB 60076 AGGCGTTGAGACAGAGCTGTGGCAACATGATGAACCCCGTCTCTAATAAATAACAAA 60135  
QY 1441 ATTTAGCTGTGTGTGTGTGACGATCTGTAAATCCAGATCTCAAGAGGCTGAGGCGGG 1500  
DB 60136 ATTTAGCTGTGTGTGTGTGACGATCTGTAAATCCAGATCTCAAGAGGCTGAGGCGGG 60195  
QY 1501 AATCTCTTGAACCTTGGAGGACAGAGGTTGCAGTGAGCCAAATCAACCAAGCACTCCA 1560

```
repeat_region      join(560. .597,3184. .3281)
                    /rpt_family="MER58"
                    /rpt_type=dispersed
repeat_region      complement(join(539. .2553,2716. .3183)))
                    /rpt_family="TTGGB81"
                    /rpt_type=dispersed
repeat_region      2558. .2581
                    /standard_name="TJ1"
                    /note="100% homology"
                    /rpt_type=direct
                    /rpt_unit=" (caggcacagtgg) 2"
repeat_region      2570. .2715
                    /rpt_family="AluSg/x"
                    /rpt_type=dispersed
repeat_region      3021. .3050
                    /standard_name="T2"
                    /note="83.3% homology"
                    /rpt_type=direct
                    /rpt_unit=" (taacaagtg) 3"
repeat_region      3283. .3558
                    /rpt_family="AluSx"
                    /rpt_type=dispersed
repeat_region      4173. .4211
                    /standard_name="TJ3"
                    /note="82.1% homology"
                    /rpt_type=direct
                    /rpt_unit=" (aaaaataattc) 3"
repeat_region      complement(4546. .4848)
                    /rpt_family="AluSx"
                    /rpt_type=dispersed
                    complement(5909. .6208)
repeat_region      /rpt_family="AluSg"
                    /rpt_type=dispersed
                    join(6332. .6573,6638. .6688)
                    /rpt_type=dispersed
```

```
/tpc_lammy-  
/mt_type=dis
```

```
/rpt_family="AluSx"
```

misc\_feature  
join(11627..11757,12434..12512,13443..13562)  
/note=">98% identity"  
/db\_xref="dbEST:AA774571"  
repeat\_region  
11936..12028  
/standard\_name="AT4"  
/note="100% homology"  
/rpt\_type=tandem  
/rpt\_unit="(attc)8"  
complement(12010..12312)  
/rpt\_family="Aluub"  
/rpt\_type=dispersed  
12327..12433  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(11946..14235)  
/rpt\_family="Alusq"  
/rpt\_type=dispersed  
15130..15159  
/note="low complexity"  
/rpt\_family="AT-rich"  
15164..15753  
/standard\_name="AT28"  
/note="82.1% homology"  
/rpt\_type=direct  
/rpt\_unit="(atgatatgctgcatatagacataac)21"  
join(15170..15289,15410..15612)  
/note="microsatellite"  
/rpt\_type=tandem  
/rpt\_unit="ta"  
15617..15731  
/note="microsatellite"  
/rpt\_type=tandem  
/rpt\_unit="cata"  
16004..16100  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(16128..16364)  
/rpt\_family="L2"  
repeat\_region  
/rpt\_type=dispersed  
complement(16407..16592)  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(16657..16889)  
/rpt\_family="MIR"  
16914..16936  
/rpt\_type=dispersed  
/note="low complexity"  
/rpt\_family="AT-rich"  
17102..17386  
/rpt\_family="Alusx"  
/rpt\_type=dispersed  
17406..17564  
/standard\_name="GA2"  
/note="81.8% homology"  
/rpt\_type=direct  
17416..17543  
/note="microsatellite"  
/rpt\_type=tandem  
/rpt\_unit="gaaa"  
17647..17702  
/rpt\_family="MIR"  
/rpt\_type=dispersed  
complement(19863..20161)  
/rpt\_family="Alusx"  
/rpt\_type=dispersed  
complement(20842..20913)  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(21366..21414)  
/note="microsatellite"  
/rpt\_type=tandem  
/rpt\_unit="gaaa"

repeat\_region 21375..21639  
/standard\_name="CT1"  
/note="75% homology"  
/rpt\_type=dispersed  
complement(21415..21536)  
/note="microsatellite"  
/rpt\_type=tandem  
/rpt\_unit="cgaa"  
repeat\_region  
complement(21524..21646)  
/note="low complexity"  
/rpt\_family="purine-rich"  
22409..22646  
/rpt\_family="L1M4"  
/rpt\_type=dispersed  
22560..22604  
/standard\_name="T4"  
/note="80% homology"  
/rpt\_type=direct  
/rpt\_unit="(ataatacaataatt)3"  
  
Query Match 99.9%; Score 1999.4; DB 8; Length 80155;  
Best Local Similarity 100.0%; Pred.No.0;  
Matches 2000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 AAAGCTTGAATAGACAAATGACCTTCAACTCTTATACCAACTCTGAGTTGGC 60  
DB 58676 AAAGCTTGAATAGACAAATGACCTTCAACTCTTATACCAACTCTGAGTTGGC 58735  
61 GACATGGCTTCTCCCTTTAGAGTCCGTGACAGCCATCTTAATAGTCATTGG 120  
DB 58736 GACATGGCTTCTCCCTTTAGAGTCCGTGACAGCCATCTTAATAGTCATTGG 58795  
121 GCCCTGATTTTAACTCTTGCTCAATTTGTTCTCTAGATGAGGCCATCAAGCT 180  
DB 58796 GCCCTGATTTTAACTCTTGCTCAATTTGTTCTCTAGATGAGGCCATCAAGCT 58855  
181 ACAGATGATCTTCAATGATGAACCCCAATGACCTCACTCAACCTTCTGAGGACC 240  
DB 58856 ACAGATGATCTTCAATGATGAACCCCAATGACCTCACTCAACCTTCTGAGGACC 58915  
241 CCTGACCGACCCGCTGGCCCTTCAATGAGCTCCCTCTGAGGACACTAC 300  
DB 58916 CCTGACCGACCCGCTGGCCCTTCAATGAGCTCCCTCTGAGGACACTAC 58975  
301 CACTGACAGGCCCCCTTCTTACCCCTATCCAGACGAAAGTCAAGCCGTATGCCA 360  
DB 58976 CACTGACAGGCCCCCTTCTTACCCCTATCCAGACGAAAGTCAAGCCGTATGCCA 59035  
361 AATCCCAACGACGCTGGGGTCTGTTGAGGGGGGATTGAGAGTGAAGCAGCTG 420  
DB 59036 AATCCCAACGACGCTGGGGTCTGTTGAGGGGGGATTGAGAGTGAAGCAGCTG 59095  
421 GGCTTGGGTCAAGTGGGACCTTGAAGAACTTTGTCTAGCTAAAGATTGTAAATG 480  
DB 59096 GGCTTGGGTCAAGTGGGACCTTGAAGAACTTTGTGTCTAGCTAAAGATTGTAAATG 59155  
481 CACCAATCAGACTCTGTCTTACGCTTAAGAGATTGTAATGCAACATCAGACTCTGTA 540  
DB 59156 CACCAATCAGACTCTGTCTTACGCTTAAGAGATTGTAATGCAACATCAGACTCTGTA 59215  
541 AATGAGCAATAGAGAGATGAGGAGGGGCAATTAAGGGGATAAAATCGGCCACCC 600  
DB 59216 AATGAGCAATAGAGAGATGAGGAGGGGCAATTAAGGGGATAAAATCGGCCACCC 59275  
601 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCAACACTGTGGAAGCTTTGTTCTTTTG 660  
DB 59276 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCAACACTGTGGAAGCTTTGTTCTTTTG 59335  
661 CTTTCAACAATAATCTGTGCTGCTCATTTCTTTGTGCCACACTACTTTATAGCTG 720  
DB 59336 CTTTCAACAATAATCTGTGCTGCTCATTTCTTTGTGCCACACTACTTTATAGCTG 59395  
721 TAACTCACTGAGGGTCTGTGCTTCAATTCCTGAAGTCAACAGACCAAGAACCACT 780

Db 59396 TAAACCTGAGGAGGTCTGTGGCTTCACTTGAAGTCAACAGACCAAGCACTCAGT 59455  
Qy 781 GGAAGGAACAAGAAATCCCCGATATGCTGCTTTAAGAGCTGTAACATCACTGCGAAGC 840  
Db 59456 GGAAGGAACAAGAAATCCCCGATATGCTGCTTTAAGAGCTGTAACATCACTGCGAAGC 59515  
Qy 841 TCTGAGCTTCACTCTGGAAGTCAAGAGCAACAACCCAGAGGAAGAACTCTG 900  
Db 59516 TCTGAGCTTCACTCTGGAAGTCAAGAGCAACAACCCAGAGGAAGAACTCTG 59575  
Qy 901 GACACACCTGAATATCTGAAGGAACAACCTCCAGACACACATCTTTCAAGAGCTGTAAC 960  
Db 59576 GACACACCTGAATATCTGAAGGAACAACCTCCAGACACACATCTTTCAAGAGCTGTAAC 59635  
Qy 961 CTCACCGCAAGAGTCTGTGGCTTCACTTTGGAAGTCAAGAGCAACAAGCAACCGGAA 1020  
Db 59636 CTCACCGCAAGAGTCTGTGGCTTCACTTTGGAAGTCAAGAGCAACAAGCAACCGGAA 59695  
Qy 1021 GGAACAATTCAGACACAGTAGAAGAAATCTGATTTTGTATCTGTGGCTTCCAGGGTTAC 1080  
Db 59696 GGAACAATTCAGACACAGTAGAAGAAATCTGATTTTGTATCTGTGGCTTCCAGGGTTAC 59755  
Qy 1081 TCCAGTCATTGAAGTCTCCATTGSCAGCTTAAAGGAAACAGAGAATGGTTGAGAGCAC 1140  
Db 59756 TCCAGTCATTGAAGTCTCCATTGSCAGCTTAAAGGAAACAGAGAATGGTTGAGAGCAC 59815  
Qy 1141 ATGTGGGAATTTGTTATGAGACAGCTTGAGATGACATAGGAGATTTGATGAAACCTTA 1200  
Db 59816 ATGTGGGAATTTGTTATGAGACAGCTTGAGATGACATAGGAGATTTGATGAAACCTTA 59875  
Qy 1201 GCTGGAAGCAGGGCCAGGAATATTAATCTAAGGAAGACATTTTGTGAGACAGTAGAT 1260  
Db 59876 GCTGGAAGCAGGGCCAGGAATATTAATCTAAGGAAGACATTTTGTGAGACAGTAGAT 59935  
Qy 1261 CTTTGACATCTGAGACATGTAATTAATCAAGCAATTAATTAAGAAATATAGCCAGTGC 1320  
Db 59936 CTTTGACATCTGAGACATGTAATTAATCAAGCAATTAATTAAGAAATATAGCCAGTGC 59995  
Qy 1321 GATGGCTCATGCTCTGTAATCCCAAGACATTTGGAGGCGCAAGGGGTGTGATCCAGAGTC 1380  
Db 59996 GATGGCTCATGCTCTGTAATCCCAAGACATTTGGAGGCGCAAGGGGTGTGATCCAGAGTC 60055  
Qy 1381 AGGCGTTGAGACAGCCTGAGCCAAACATGGTGAACCCCGTCTCTACTATAAAATATCAAA 1440  
Db 60056 AGGCGTTGAGACAGCCTGAGCCAAACATGGTGAACCCCGTCTCTACTATAAAATATCAAA 60115  
Qy 1441 ATTAGCCTGTGTGTGTGAGACGATCTGTATCCAGTACTCAGAGGCTGAGGCAAGG 1500  
Db 60116 ATTAGCCTGTGTGTGTGAGACGATCTGTATCCAGTACTCAGAGGCTGAGGCAAGG 60175  
Qy 1501 AATCTCTTGAACCTTGGAGGCAAGGTTGAGAGTCCAGATACACACAGACCTCCA 1560  
Db 60176 AATCTCTTGAACCTTGGAGGCAAGGTTGAGAGTCCAGATACACACAGACCTCCA 60235  
Qy 1561 TCCGTGGTGAAGAGGAGACTGTCTCAAAAAAAGAAAAAAGAAAAAGAAAAAT 1620  
Db 60236 TCCGTGGTGAAGAGGAGACTGTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAT 60295  
Qy 1621 ATATCAAGAAATATGACAGTAAACATTTATTAACATTAATGACACAGGCAATACA 1680  
Db 60296 ATATCAAGAAATATGACAGTAAACATTTATTAACATTAATGACACAGGCAATACA 60355  
Qy 1681 CTAAAGTTTTTACATGATTAACATTTAATCTTAACAATAGCCCTTAAGAGTCAAGTGC 1740  
Db 60356 CTAAAGTTTTTACATGATTAACATTTAATCTTAACAATAGCCCTTAAGAGTCAAGTGC 60415  
Qy 1741 TGTATTAATCCACTTATATGATTAAGAACTGAAGTACAGAAAGGTCAAGTAGAGAA 1800  
Db 60416 TGTATTAATCCACTTATATGATTAAGAACTGAAGTACAGAAAGGTCAAGTAGAGAA 60475  
Qy 1801 TGCCCATGCTTGAATCTCACTTTTGAAGCAAGTGTACAGAAATCTGTGTAGAAAT 1860

Db 60476 TGCCCATGCTTGCATCTTCAGATTTTGAAGCAACTGTTACAGAAATCTGTGTAGAAAT 60535  
Qy 1861 GCTTAACAAGATGAGTCAAGGGTGTGGAGTACTAGTCTGATTTGGCAAGTTGGG 1920  
Db 60536 GCTTAACAAGATGAGTCAAGGGTGTGGAGTACTAGTCTGATTTGGCAAGTTGGG 60595  
Qy 1921 ATGAAGATGATGAAGACAGCTTTGACAGAAAGCTGACACTTGGCAACTCTGTGGGA 1980  
Db 60596 ATGAAGATGATGAAGACAGCTTTGACAGAAAGCTGACACTTGGCAACTCTGTGGGA 60655  
Qy 1981 CCTGAAGGGTTAGAGGACT 2001  
Db 60656 CCTGAAGGGTTAGAGGACT 60676

RESULT 6  
AF222854 80202 bp DNA linear PRI 17-JUL-2000  
LOCUS AF222854  
DEFINITION Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.  
ACCESSION AF222854  
VERSION AF222854.1 GI:9246844  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE 1 (bases 1 to 80202)  
Barry,A.E., Bateman,M., Howman,E.V., Cancellia,M.R., Tainton,K.M.,  
Irvine,D.V., Saffery,R. and Choo,K.H.  
The 10q25 neocentromere and its inactive progenitor have identical  
primary nucleotide sequence: further evidence for epigenetic  
modification  
Genome Res. 10 (6), 832-838 (2000)  
JOURNAL PUBMED 10854414  
REFERENCE 2 (bases 1 to 80202)  
Barry,A.E.  
Direct Submissiion  
Submitted (11-JAN-2000) Chromosome Research Unit, The Murdoch  
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,  
Parkville, Melbourne, Victoria 3052, Australia  
Human genomic sequence from 10q25.2, clone1b=NC, second release  
REMARK  
FEATURES  
source  
1. 80202  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10q25.2"  
/clone="NC"  
complement(598..3183)  
/rpt\_family="Tigger1"  
2587..2715  
/rpt\_family="Alu"  
3283..3553  
/rpt\_family="Alu"  
repeat\_region  
/rpt\_family="Alu"  
complement(4548..4809)  
/rpt\_family="Alu"  
complement(5919..6211)  
/rpt\_family="Alu"  
6257..6599  
/rpt\_family="MLT2B2"  
complement(8848..9123)  
/rpt\_family="Alu"  
repeat\_region  
complement(10913..11198)  
/rpt\_family="Alu"  
complement(12062..12313)  
/rpt\_family="Alu"  
complement(13957..14236)  
/rpt\_family="Alu"  
17133..17424  
/rpt\_family="Alu"  
complement(19882..20197)  
repeat\_region



```
/rpt_family="Alu"
repeat_region complement(22664, .22956)
/rpt_family="Alu"
repeat_region complement(23935, .24113)
/rpt_family="MIR"
repeat_region 24585, .24884
/rpt_family="Alu"
repeat_region complement(25618, .25973)
/rpt_family="THE1"
repeat_region 26732, .26970
/rpt_family="MER2"
repeat_region 27173, .27449
/rpt_family="Alu"
repeat_region 27526, .27796
/rpt_family="Alu"
repeat_region 28702, .28978
/rpt_family="Alu"
repeat_region complement(30668, .30736)
/rpt_family="MLT1"
repeat_region complement(31460, .31993)
/rpt_family="MER1"
repeat_region complement(32130, .32248)
/rpt_family="MLT1"
repeat_region 32685, .32961
/rpt_family="Alu"
repeat_region 33113, .33403
/rpt_family="Alu"
repeat_region complement(34169, .34233)
/rpt_family="MIR"
repeat_region 34332, .34639
/rpt_family="Alu"
repeat_region complement(35993, .36081)
/rpt_family="MIR"
repeat_region complement(38438, .38721)
/rpt_family="Alu"
repeat_region 39081, .39333
/rpt_family="Alu"
repeat_region complement(42817, .42877)
/rpt_family="MIR"
repeat_region complement(44845, .45119)
/rpt_family="Alu"
repeat_region 45532, .45665
/rpt_family="MIR"
repeat_region complement(46257, .46490)
/rpt_family="Alu"
repeat_region 48262, .48578
/rpt_family="Alu"
repeat_region complement(48687, .50337)
/rpt_family="L1"
repeat_region complement(50349, .50611)
/rpt_family="Alu"
repeat_region complement(50670, .50831)
/rpt_family="L1"
repeat_region complement(51085, .51374)
/rpt_family="Alu"
repeat_region complement(52905, .53197)
/rpt_family="Alu"
repeat_region 53896, .54467
/rpt_family="LTR12"
repeat_region 59129, .59719
/rpt_family="LTR12"
repeat_region 60018, .60302
/rpt_family="Alu"
repeat_region complement(60338, .60470)
/rpt_family="MIR"
repeat_region complement(61351, .61521)
/rpt_family="MERS"
repeat_region 63135, .63416
/rpt_family="Alu"
repeat_region 63724, .63889
/rpt_family="MERS"
repeat_region 64107, .64182
/rpt_family="MIR"
```

```
repeat_region 65459, .65548
/rpt_family="MERS"
repeat_region 66439, .66618
/rpt_family="MER20"
repeat_region complement(67562, .68084)
/rpt_family="MER41"
repeat_region complement(72861, .73139)
/rpt_family="Alu"
repeat_region 73756, .74037
/rpt_family="Alu"
repeat_region 75588, .75853
/rpt_family="Alu"
repeat_region complement(77441, .77729)
/rpt_family="Alu"

ORIGIN
Query Match 99.9%; Score 1999.4; DB 8; Length 80202;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAGGCTTCGAAATCAGACAAATGCCTTCAAACTCTTATACCAACTCTGAGTTGGG 60
DB 58696 AAAGGCTTCGAAATCAGACAAATGCCTTCAAACTCTTATACCAACTCTGAGTTGGG 58755
QY 61 GACATGGCTTTCCTCCCTTCTAGAGTCTGTGACAGCCATCTTGTATAGTCGATTTGG 120
DB 58756 GACATGGCTTTCCTCCCTTCTAGAGTCTGTGACAGCCATCTTGTATAGTCGATTTGG 58815
QY 121 GCCCTGTATTTTAACTCTTGTGTCAAATTTGTTTCTTAGAGATCGAGGCATCAAGCT 180
DB 58816 GCCCTGTATTTTAACTCTTGTGTCAAATTTGTTTCTTAGAGATCGAGGCATCAAGCT 58875
QY 181 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCACTAACAACCTTCTGAGAGACC 240
DB 58876 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCACTAACAACCTTCTGAGAGACC 58935
QY 241 CCTGACGACGACCGGCGGCTTCAATGAGCCCTTAAGAGCTCCCTGAGAGCACTAC 300
DB 58936 CCTGACGACGACCGGCGGCTTCAATGAGCCCTTAAGAGCTCCCTGAGAGCACTAC 58995
QY 301 CACTGACGAGGCGCCCTTCTTACCCCTATCAGACGAGAGTAGCTACAGCGGTATGCCA 360
DB 58996 CACTGACGAGGCGCCCTTCTTACCCCTATCAGACGAGAGTAGCTACAGCGGTATGCCA 59055
QY 361 AATCCCAACAGACGCTGGGCTGTCTGTTTGAAGGGGGAATTGAAGGTGAACGACGCTG 420
DB 59056 AATCCCAACAGACGCTGGGCTGTCTGTTTGAAGGGGGAATTGAAGGTGAACGACGCTG 59115
QY 421 GGCTTCGGGTCAAGTGGGGAAGTGGAGAACTTTGTGTAGCTAAAGGATGTAAAG 480
DB 59116 GGCTTCGGGTCAAGTGGGGAAGTGGAGAACTTTGTGTAGCTAAAGGATGTAAAG 59175
QY 481 CACCAATCAGACCTGTGTCTAGCTAAAGATTTGTAATGCAACATGACACTCTGTGA 540
DB 59176 CACCAATCAGACCTGTGTCTAGCTAAAGATTTGTAATGCAACATGACACTCTGTGA 59235
QY 541 AATGACCAATCAGACGATGTGGCGGGGTCAAAATTAAGGATGAATAAATCGGCCAACC 600
DB 59236 AATGACCAATCAGACGATGTGGCGGGGTCAAAATTAAGGATGAATAAATCGGCCAACC 59295
QY 601 GAGCCAGAGTGGCAACCACTCGGCTCCCTTCCACACTGTGGAAGCTTGTCTTTTG 660
DB 59296 GAGCCAGAGTGGCAACCACTCGGCTCCCTTCCACACTGTGGAAGCTTGTCTTTTG 59355
QY 661 CTCTTCACATTAATCTTGTCTGTCTATCTTATTTGTGTGCAACACTTATATGAGCTG 720
DB 59356 CTCTTCACATTAATCTTGTCTGTCTATCTTATTTGTGTGCAACACTTATATGAGCTG 59415
QY 721 TAACTCTCACTGGAGGGGTCTGTGCTTATCTGGAAGTCAACAGACCAACCACT 780
DB 59416 TAACTCTCACTGGAGGGGTCTGTGCTTATCTGGAAGTCAACAGACCAACCACT 59475
QY 781 GGAAGGAACAAAGAACTCCGATGTGTCTCTTAAAGACTGTAACTCACTGCGAAGC 840
```

Db 59476 GGAAGAAACAAGAACTCCGATGTGCTCTTTAAGAGCTGTAACACTACCTGCGCAAC 59535  
 QY 841 TCTGCAGCTTCACTCCTGAAAGTCAGTACCAACCAACCCAGAGAAAGAACTCTG 900  
 Db 59536 TCTGCAGCTTCACTCCTGAAAGTCAGTACCAACCAACCCAGAGAAAGAACTCTG 59595  
 QY 901 GACACACTGAAATATCTGAAAGAACAACTCCAGACACACATCTTTCAGAGCTGTACA 960  
 Db 59596 GACACACTGAAATATCTGAAAGAACAACTCCAGACACACATCTTTCAGAGCTGTACA 59655  
 QY 961 CTCACCGCAAGGCTCTGAGCTTCACTTCTTGAAGTCAGACAGACCAAGAACCCACCGGAA 1020  
 Db 59656 CTCACCGCAAGGCTCTGAGCTTCACTTCTTGAAGTCAGACAGACCAAGAACCCACCGGAA 59715  
 QY 1021 GGAAACAATTTCCAGACACAGTGAAGAAATCTGTATTTTGAATCTGTGAGCTTCCAGGGTTAC 1080  
 Db 59716 GGAAACAATTTCCAGACACAGTGAAGAAATCTGTATTTTGAATCTGTGAGCTTCCAGGGTTAC 59775  
 QY 1081 TCCAGTCATTGAAGTCTCCATTCAGCTTAAAGAAACAGAGAAATGGTTTGAAGAGACAC 1140  
 Db 59776 TCCAGTCATTGAAGTCTCCATTCAGCTTAAAGAAACAGAGAAATGGTTTGAAGAGACAC 59835  
 QY 1141 ATGTGGAAATTTTATGAGACAGGCTTGAGATGCAATAGGCAATTTCTGATCAAACTTA 1200  
 Db 59836 ATGTGGAAATTTTATGAGACAGGCTTGAGATGCAATAGGCAATTTCTGATCAAACTTA 59895  
 QY 1201 GCTGGAAGACGGCCGAGAAATTAATCTTAAGAAAGCACTTTTGTAGACATAGTAGT 1260  
 Db 59896 GCTGGAAGACGGCCGAGAAATTAATCTTAAGAAAGCACTTTTGTAGACATAGTAGT 59955  
 QY 1261 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAAGAAAAATATAGCCAGGTGC 1320  
 Db 59956 CTTTGCATCTGAGACATGTAGATTATCAAGCAATTAATTAAGAAAAATATAGCCAGGTGC 60015  
 QY 1321 GATGCTCATGCTGTATATCCAGACCTTTTGGAGGCCAAGGGTGTGATCACAGGTGC 1380  
 Db 60016 GATGCTCATGCTGTATATCCAGACCTTTTGGAGGCCAAGGGTGTGATCACAGGTGC 60075  
 QY 1381 AGGCGTTGACAGACAGCTGTGCCCAACATGTTGAAACCCCGTCTCTACTTAAATAATACAAA 1440  
 Db 60076 AGGCGTTGACAGACAGCTGTGCCCAACATGTTGAAACCCCGTCTCTACTTAAATAATACAAA 60135  
 QY 1441 ATTAGCCTGCTGTGTGGACGCACTCTGTATCCAGTACTCAGAGAGCTGAGGACAGGG 1500  
 Db 60136 ATTAGCCTGCTGTGTGGACGCACTCTGTATCCAGTACTCAGAGAGCTGAGGACAGGG 60195  
 QY 1501 AATCTCTTGAACCTTGGAGGACAGAGGTTGACAGTGAAGTCAACACAGCACTCCA 1560  
 Db 60196 AATCTCTTGAACCTTGGAGGACAGAGGTTGACAGTGAAGTCAACACAGCACTCCA 60255  
 QY 1561 TCTCGGTGACAGAGGACAGTCTGTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGT 1620  
 Db 60256 TCTCGGTGACAGAGGACAGTCTGTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGT 60315  
 QY 1621 ATATCAAGAAATATGACAGGTATCACTTATTAACAATTACTATGACACAGCAATACA 1680  
 Db 60316 ATATCAAGAAATATGACAGGTATCACTTATTAACAATTACTATGACACAGCAATACA 60375  
 QY 1681 CTAAGTCTTTTACATGATTAATCACTTATTAATCTTAAACATAGCCCTATGAGTCAAGTGC 1740  
 Db 60376 CTAAGTCTTTTACATGATTAATCACTTATTAATCTTAAACATAGCCCTATGAGTCAAGTGC 60435  
 QY 1741 TGTATTATATCTCACTTATTAATGATTAAGAAACAGAAAGTCAAGTGAAGAAA 1800  
 Db 60436 TGTATTATATCTCACTTATTAATGATTAAGAAACAGAAAGTCAAGTGAAGAAA 60495  
 QY 1801 TGGCCATGCTTGCATCTCACTTTTGAAGCACTTTACAGAAATCTGTGTAGAAAT 1860  
 Db 60496 TGGCCATGCTTGCATCTCACTTTTGAAGCACTTTACAGAAATCTGTGTAGAAAT 60555  
 QY 1861 GCTCTAACAAATGTAGTCAAGGGGTTGGAAGTACTGAGCTGAGTGGCAGTTGGGG 1920

Db 60556 GCTCTAACAAATGTAGTCAAGGGGTTGGAGAGTACTAGTCTGAGTTGGGAGTTGGGG 60615  
 QY 1921 ATGGAAGATGATGATGAAGAAACAGCTTGAACAGAAAGCTGACACTTGTGGCA 1980  
 Db 60616 ATGGAAGATGATGATGAAGAAACAGCTTGAACAGAAAGCTGACACTTGTGGCA 60675  
 QY 1981 CCTTGAAGGTTTGAAGGACT 2001  
 Db 60676 CCTTGAAGGTTTGAAGGACT 60696  
 RESULT 7  
 AL355530  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone RP11-338L11 on chromosome 10 contains the 5' end of the gene for novel gene (K1A0534) and a CpG island, complete sequence.  
 ACCESSION  
 AL355530  
 VERSION  
 AL355530.6 GI:8894326  
 KEYWORDS  
 HTG; KIA0534.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 1 (bases 1 to 196934)  
 Bird.C.  
 Direct Submission  
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
 Clone requests: clonerequests@sanger.ac.uk  
 On Jul 1, 2000 this sequence version replaced gi:8670617.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 -----Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Genet. code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: vegas@sanger.ac.uk  
 -----  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 RP11-338L11 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBac3.6.  
 FEATURES  
 source  
 Location/Qualifiers  
 1. 196934  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-338L11"  
 /clone\_1ib="RP11-11.2"  
 1  
 /note="Clone left end: RP11-338L11"  
 join(68829..\_69507,95654..\_95737,97168..\_97281,  
 103062..\_103190,104794..\_105002,135506..\_135680,  
 141023..\_141110,146500..\_146637)  
 /gene="RP11-537G20.1"  
 /locus\_tag="RP11-537G20.1..002"



Db 17498 CACTCAGAGGCCCCCTTCTTCAACCCCTATCCAGACAGGAAGTACTACAGGGCTCATCGCA 17557  
Qy 361 AATCCCAACAGAGCTGGGGGTCTCTTTTGGAGGGGGATTTGAGAGTGAAAGCCAGCTG 420  
Db 17558 AATCCCAACAGAGCTGGGGGTCTCTTTTGGAGGGGGATTTGAGAGTGAAAGCCAGCTG 17617  
Qy 421 GGCTTCTGGGTCAAGGTGGGGACTTGGAGAACTTTGTGTCTAGCTAAAGATTTGTAATG 480  
Db 17618 GGCTTCTGGGTCAAGGTGGGGACTTGGAGAACTTTGTGTCTAGCTAAAGATTTGTAATG 17677  
Qy 481 CACCAATCAGACTCTGTCTAGCTAAAGATTTGTAATGCAACCAATCAGACTCTGTGTA 540  
Db 17678 CACCAATCAGACTCTGTCTAGCTAAAGATTTGTAATGCAACCAATCAGACTCTGTGTA 17737  
Qy 541 AATAGGACCAATCAGAGGATGTGGCGGGTCAAAATTAAGAGATTAATACTGGCCACC 600  
Db 17738 AATAGGACCAATCAGAGGATGTGGCGGGTCAAAATTAAGAGATTAATACTGGCCACC 17797  
Qy 601 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCCACACTGGGAAGCTTTGTCTTTG 660  
Db 17798 GAGCCAGAGTGGCAACCACTCGGGTCCCTTCCACACTGGGAAGCTTTGTCTTTG 17857  
Qy 661 CTCTTCACAATAAATCTGTGCTGCTCACTTTGTGTCAACTACCTTTATGAGCTG 720  
Db 17858 CTCTTCACAATAAATCTGTGCTGCTCACTTTGTGTCAACTACCTTTATGAGCTG 17917  
Qy 721 TAACTCTACCTGGAAGGTCTGTGGCTTCACTCTGAAGTCAACAGACCAAGCCACT 780  
Db 17918 TAACTCTACCTGGAAGGTCTGTGGCTTCACTCTGAAGTCAACAGACCAAGCCACT 17977  
Qy 781 GGAGGAAACAAAGAACTCCCGATGTGTCTGCTTTAAGCTGAACACTCACTGGAGAG 840  
Db 17978 GGAGGAAACAAAGAACTCCCGATGTGTCTGCTTTAAGCTGAACACTCACTGGAGAG 18037  
Qy 841 TCTGAGCTTCACTCTGGAAGTCACTGAGACCAAAACCACAGAGAAAGAACTCTG 900  
Db 18038 TCTGAGCTTCACTCTGGAAGTCACTGAGACCAAAACCACAGAGAAAGAACTCTG 18097  
Qy 901 GACACACTGAATATCTGAAGAACAAACTCCAGACACACATCTTTCAGAGCTGTACA 960  
Db 18098 GACACACTGAATATCTGAAGAACAAACTCCAGACACACATCTTTCAGAGCTGTACA 18157  
Qy 961 CTGACCGCAAGGCTCTGTGCTTCACTTTGAAGTCAAGCAACCAAGCCACCGGAA 1020  
Db 18158 CTGACCGCAAGGCTCTGTGCTTCACTTTGAAGTCAAGCAACCAAGCCACCGGAA 18217  
Qy 1021 GGAACAAATTCAGACACAGTAGGAATCTGTATTTTGTGATCTGTGCTTTCAGGGTTAC 1080  
Db 18218 GGAACAAATTCAGACACAGTAGGAATCTGTATTTTGTGATCTGTGCTTTCAGGGTTAC 18277  
Qy 1081 TCCAGTCAATTGAAGTCTCATTTGCAAGCTTAAAGAAACAGAAATGTTTGGAGAGCAC 1140  
Db 18278 TCCAGTCAATTGAAGTCTCATTTGCAAGCTTAAAGAAACAGAAATGTTTGGAGAGCAC 18337  
Qy 1141 ATGTGGGAATTTGTAATGACAGGCTTGAATGACATAGGGCAATTTCTGATCAAACTTA 1200  
Db 18338 ATGTGGGAATTTGTAATGACAGGCTTGAATGACATAGGGCAATTTCTGATCAAACTTA 18397  
Qy 1201 GCTGGAAGCAAGGCGCAGGAATAATATCTAAGAAACAGTTTGTGACAGTAGTAGT 1260  
Db 18398 GCTGGAAGCAAGGCGCAGGAATAATATCTAAGAAACAGTTTGTGACAGTAGTAGT 18457  
Qy 1261 CTTTGCACTGAGACATGATGATTAAGCAATTAATTAGAAAAATATAGCCAGGTGC 1320  
Db 18458 CTTTGCACTGAGACATGATGATTAATCAAGCAATTAATTAGAAAAATATAGCCAGGTGC 18517  
Qy 1321 GATGGCTCATGCTGTGTAATCCAGACCTTTGGAGGCCAAGGGGTGTGATCAAGAGTTC 1380  
Db 18518 GATGGCTCATGCTGTGTAATCCAGACCTTTGGAGGCCAAGGGGTGTGATCAAGAGTTC 18577  
Qy 1381 AGGGCTTCAGACCAAGCTGGCCAAACATGTGAAACCCCGCTCTACTAAATAACAAA 1440

Db 18578 AGGGCTTCAGACCAAGCTGGCCAAACATGTGAAACCCCGTCTCTACTTAAAAATACAAA 18637  
Qy 1441 ATTAGCTGTGTGTGSCAGCATCTGTAATCCAGTACTCAGAGGCTGAGCAGGG 1500  
Db 18638 ATTAGCTGTGTGTGSCAGCATCTGTAATCCAGTACTCAGAGGCTGAGCAGGG 18697  
Qy 1501 AATCTTTGAACTTGGAGGAGAGTTCAGTGAAGCAATGACACACACACTCA 1560  
Db 18698 AATCTTTGAACTTGGAGGAGAGTTCAGTGAAGCAATGACACACACACTCA 18757  
Qy 1561 TCTGGGTGACAGACCGGACCTCTCTCAAAAAAAGGAAAGGAAAT 1620  
Db 18758 TCTGGGTGACAGACCGGACCTCTCTCAAAAAAAGGAAAGGAAAT 18817  
Qy 1621 ATAAATCAAGAAATATGACAGGTAACTTAATCAACCTTATGACACAGCAATACA 1680  
Db 18818 ATAAATCAAGAAATATGACAGGTAACTTAATCAACCTTATGACACAGCAATACA 18877  
Qy 1681 CTAAAGTGTATTAATCATGATTAATCTCAATTAATCTTAACAAATGCTTATGAAGTGC 1740  
Db 18878 CTAAAGTGTATTAATCATGATTAATCTCAATTAATCTTAACAAATGCTTATGAAGTGC 18937  
Qy 1741 TGTATTAATCTCACTTTATGATTAAGAAAGTAAAGTAAAGTCAAGAGAA 1800  
Db 18938 TGTATTAATCTCACTTTATGATTAAGAAAGTAAAGTAAAGTCAAGAGAA 18997  
Qy 1801 TGGCCATGCTTCACTTCTCACTTTTGAAGCAACTGTTACAGAAATCTGTGTGAAT 1860  
Db 18998 TGGCCATGCTTCACTTCTCACTTTTGAAGCAACTGTTACAGAAATCTGTGTGAAT 19057  
Qy 1861 GCTTAAACAAATGTGAGTCAAGGGGTGGAGATCTGAGTCTGAGTGGCACTTGGG 1920  
Db 19058 GCTTAAACAAATGTGAGTCAAGGGGTGGAGATCTGAGTCTGAGTGGCACTTGGG 19117  
Qy 1921 ATGAAGATGATGAAGAAACGCTTGAACAGAAAGCTGACCTTGGCACTGTGGGA 1980  
Db 19118 ATGAAGATGATGAAGAAACGCTTGAACAGAAAGCTGACCTTGGCACTGTGGGA 19177  
Qy 1981 CCTTGAAGGTTAGAGGACT 2001  
Db 19178 CCTTGAAGGTTAGAGGACT 19198

RESULT 8  
AC093818/c 46324 bp DNA linear PRI 21-APR-2005  
LOCUS DEFINITION Homo sapiens BAC clone RP11-366L15 from 2, complete sequence.  
ACCESSION AC093818 AC020681  
VERSION AC093818.3 GI:18497233  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 46324)  
Kozlowski, A., Doeber, A. and Nguyen, C.  
The sequence of Homo sapiens BAC clone RP11-366L15  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 46324)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 46324)  
REFERENCE 3 (bases 1 to 46324)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 46324)  
REFERENCE 4 (bases 1 to 46324)  
AUTHORS Waterston, R.

TITLE Direct Submission  
JOURNAL Submitted (08-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 46324)  
AUTHORS Watson, R.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 6 (bases 1 to 46324)  
AUTHORS Wilson, R.K.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
On Feb 5, 2002 this sequence version replaced gi:15809168.

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0366L15  
Drafting Center: WIBR

-----  
NOTICE:  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPci-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Cacanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is RP11-343H10. Actual start of this clone is at base position 1 of RP11-366L15.

The clone RP11-366L15 contains a 150 bp insertion not sequenced in the right neighboring clone, RP11-343H10. This insertion sequence is being submitted as H\_NH0366L15\_F1.

The sequence of AC020681 has been incorporated into AC093818.

FEATURES  
source Location/Qualifiers  
1..46324  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-366L15"  
/clone\_lib="RPci-11"  
4019..5261  
/note="CpG island (%GC=69.8, o/e=0.89, #CpGs=143)"  
gene  
4654..41594  
/gene="PDK1"  
join(4654..4892,7254..7395,10766..10837,13049..13233,

13524..13619,15402..15479,17287..17363,19272..19370,34824..34934,41481..41594)  
/gene="PDK1"  
join(4697..4892,7254..7395,10766..10837,13049..13233,13524..13619,15402..15479,17287..17363,19272..19370,34824..34934,41481..41594)  
/gene="PDK1"  
/note="Homo sapiens pyruvate dehydrogenase kinase, isoenzyme 1, mRNA (cDNA clone MGC:24867 IMAGE:4778360), complete cds.; H\_NH0366L15.1  
This gene was based on gi(24660127)  
Continues as H\_NH0343H10.1"  
/codon start=1  
/product="unknown"  
/protein\_id="AAV14915.1"  
/db\_xref="GI:62822362"  
/translation="MRRLALNGAALAGPGLRAAGSRSSDSGSSPASRGVVGQVDFARFSPSLSKMOKFLDPSVNAKETSMPFROELVRLANIMKEISLIPMLRTPSVOLQSWYIOSLELDPEKDSAEADKA1YPTDVIIRRNHNVIIPMAQVIEKSKFGVDPVTSQNOVQYELDRFYSRISIMLINOHSLRFGKKGKSPSRKHGSIINPCNVLAVIDGYENARRLCDLIYINSPELEELNAKSPQPIQVYVPSHLYTMVELRKMARATMEHHANRGVTP1QVHTVLENEDELTVMSDRGGGVPLRKIDRLFNYMYSTARPRVETSRVAVPLAGFGYGLPSRLYAYVQGDILKYSLBGYDAVIYIK"

## ORIGIN

Query Match 41.4%; Score 828; DB 8; Length 46324;  
Best Local Similarity 88.3%; Pred. No. 2,4e-209;  
Matches 922; Conservative 0; Mismatches 120; Indels 2; Gaps 2;

QY 1 AAAGGCTTGAATACAGACATGCTTCAACTCTTTATCAACCTCTGAGTTGGG 60  
DB 39867 AAAGCTTCGAATACAGACATGCTTCAACTCTTTATCAACCTCTGAGTTGGG 39808  
QY 61 GACATGGCTTCCCTTTCTAGGTCCTGTGAAGCATCTTGTAATAGTCGATTGG 120  
DB 39807 AACATGGCTTCCCTTTCTAGGTCCTGTGAAGCATCTTGTAATAGTCGATTGG 39748  
QY 121 GCCCTGATTTTAACTTGTGCAATTTGTTCTTAGAGTCGAGCCATCAAGCT 180  
DB 39747 ACCCTGATTTTAACTTGTGCAATTTGTTCTTAGAGTCGAGCCATCAAGCT 39688  
QY 181 ACAGATGATCTTACAAATGATACCCCAATGAGTCTCACTTAACCTTGCTGAGGACC 240  
DB 39687 ATAGATGATCTTACAAATGATACCCCAATGAGTCTCACTTAACCTTGCTGAGGACC 39628  
QY 241 CCTGACCGACCCGCTGGCTTTCAATGAGCTTCCCTCTGAGAGCACTAC 300  
DB 39627 CCTGACCGACCCGCTGGCTTTCAATGAGCTTCCCTCTGAGAGCACTAC 39568  
QY 301 CACTGAGGAGCCCTTTCTACCCCTATTCAGAGAGATGACTACGCGTCAATG-CC 359  
DB 39567 AACTGAGGAGCCCTTTCTACCCCTATTCAGAGAGATGACTACGCGTCAATG-CC 39508  
QY 360 AATATCCCAACAGACGTGGGAGTCTGTTTGAAGGGGAGTTGAGGTGAAGCCAGCT 419  
DB 39507 AATATCCCAACAGACGTGGGAGTCTGTTTGAAGGGGAGTTGAGGTGAAGCCAGCT 39448  
QY 420 GGGCTTCTGGGTCAGGTGGGAGCTTGAAGACTTTTGTCTTACGTTAAAGATTGTAAT 479  
DB 39447 GGGCTTCTGGGTCAGGTGGGAGCTTGAAGACTTTTGTCTTACGTTAAAGATTGTAAT 39388  
QY 480 GCACCAATCAGACCTGTGTCTTACGTTAAAGATTGTAATGACCAATCAGACCTGT 539  
DB 39387 GCACCAATCAGACCTGTGTCTTACGTTAAAGATTGTAATGACCAATCAGACCTGT 39328  
QY 540 AAAATGACCAATCAGACGATGGGCGGGGCTCAATTAAGGAGATTAATAATCTGGCCACC 599  
DB 39327 AAAATGACCAATCAGACGATGGGCGGGGCTCAATTAAGGAGATTAATAATCTGGCCACC 39268  
QY 600 CGAGCAGACGATGGCAACCACTGGGCTCCCTTCCACACTGTGGAAGCTTGTCTTTT 659  
DB 39267 CGAGCAGACGATGGCAACCACTGGGCTCCCTTCCACACTGTGGAAGCTTGTCTTTT 39208



```

/rpt_family="L1"
repeat_region 3370..3614
/rpt_family="L1"
repeat_region 3619..4195
/rpt_family="L1"
repeat_region 4790..5239
/rpt_family="ERVU"
repeat_region 5697..5736
/rpt_family="(GA)n"
repeat_region 6172..6629
/rpt_family="L1"
repeat_region 6300..6662
/rpt_family="(TA)n"
repeat_region 6664..6698
/rpt_family="(TG)n"
repeat_region 6699..6805
/rpt_family="L1"
repeat_region 6806..7175
/rpt_family="MALR"
repeat_region 7178..8788
/rpt_family="MALR"
repeat_region 8789..9159
/rpt_family="MALR"
repeat_region 9161..9573
/rpt_family="L1"
repeat_region 9971..10070
/rpt_family="Alu"
repeat_region 10094..10263
/rpt_family="MALR"
repeat_region 10420..10461
/rpt_family="AT_rich"
repeat_region 10764..10881
/rpt_family="L2"
repeat_region 10882..11191
/rpt_family="Alu"
repeat_region 11197..11391
/rpt_family="Alu"
repeat_region 11392..11828
/rpt_family="L2"
repeat_region 11851..11954
/rpt_family="ERVU"
repeat_region 12429..13851
/rpt_family="L1"
repeat_region 14181..14202
/rpt_family="AT_rich"
repeat_region 15058..15223
/rpt_family="MER1_type"
repeat_region 15238..15523
/rpt_family="Alu"
repeat_region 16058..16320
/rpt_family="Alu"
repeat_region 16328..16758
/rpt_family="ERVK"
repeat_region 16896..17262
/rpt_family="L1"
repeat_region 17405..17867
/rpt_family="L1"
repeat_region 20694..21124
/rpt_family="ERVU"
repeat_region 21138..21261
/rpt_family="Achobo"
repeat_region 22074..22304
/rpt_family="MER2_type"
repeat_region 22381..22732
/rpt_family="L1"
repeat_region 23389..23835
/rpt_family="ERVU"
repeat_region 24144..24174
/rpt_family="(TA)n"
repeat_region 26614..26822
/rpt_family="MIR"
repeat_region 27691..28064
/rpt_family="MALR"

```

```

repeat_region 28148..28260
/rpt_family="L1"
repeat_region 28280..28373
/rpt_family="CT-rich"
repeat_region 28416..28528
/rpt_family="L2"
repeat_region 29982..30048
/rpt_family="L2"
repeat_region 30453..30778
/rpt_family="ERVU"
repeat_region 30782..30902
/rpt_family="L1"

```

Query Match 40.5%; Score 810.6; DB 8; Length 131458;  
 Best Local Similarity 87.7%; Pred. No. 9.9e-205;  
 Matches 938; Conservative 0; Mismatches 109; Indels 22; Gaps 4;

```

QY 1 AAGGCTTCGAATGAGCAATGACCTTTCATCTTATACCACTCTGAGTTGGC 60
Db 73714 AAGGCTTCGAATGAGCAATGACCTTTCATCTTATACCACTCTGAGTTGGC 73773
QY 61 GACATGAGCTTCCCTTCTAGAGTCTGTGACAGCCATCTGTAATAGTGCATTGG 120
Db 73774 AATGAGCTTCCCTTCTAGAGTCTGTGACAGCCATCTGTAATAGTGCATTGG 73833
QY 121 GCCCTGATTTTAACTCTGTGCAATTTGTTCTCTAGAGTGAAGGCCATCAAGT 180
Db 73834 GCCCTGATTTTAACTCTGTGCAATTTGTTCTCTAGAGTGAAGGCCATCAAGT 73893
QY 181 ACAGATGATCTTCAATGATGAACCCCAATGAGCTCACTAACACTTGTGAGGACC 240
Db 73894 ACAGATGATCTTCAATGATGAACCCCAATGAGCTCACTAACACTTGTGAGGACC 73953
QY 241 CTTGACGACCGCTGCGCTTTCAATGAGCTTCAATGAGCTTCAATGAGCTTCA 300
Db 73954 CTTGACGACCGCTGCGCTTTCAATGAGCTTCAATGAGCTTCAATGAGCTTCA 74013
QY 301 CACTGACGAGCCCTTCTTCACTCTTATCAGAGGAATGATGACGAGCTATG-CC 359
Db 74014 AACTGACGAGCCCTTCTTCACTCTTATCAGAGGAATGATGACGAGCTATG-CC 74073
QY 360 AATCCCAACAGAGCTGAGGAGTCTGTTTGAAGGAGGATTAAGAGGAGGAGCC 419
Db 74074 AATCCCAACAGAGCTGAGGAGTCTGTTTGAAGGAGGATTAAGAGGAGGAGCC 74133
QY 420 GGGCTTCTGGTCAAGTGGGAGCTTGAAGATTTTGTCTAGCTAAAGATTGAA 479
Db 74134 GAGCTTCTGGTCAAGTGGGAGCTTGAAGATTTTGTCTAGCTAAAGATTGAA 74193
QY 480 GCACCAATCAGACTTGTGTCTAGCTAAAGATTGAAATGACCAATCAGACTCTGT 539
Db 74194 GCACCAATCAGACTTGTGTCTAGCTAAAGATTGAAATGACCAATCAGACTCTGT 74237
QY 540 AAAATGACCAATCAGAGATGAGGAGGAGGCAATTAAGGAGTAAGAACTG-CC 595
Db 74238 AAAATGACCAATCAGAGATGAGGAGGAGGCAATTAAGGAGTAAGAACTG-CC 74297
QY 596 CACCCGAGCAGAGTGGACCACTCGAGTCCCTTCACTGAGAGGCTTTGTC 655
Db 74298 CACCCGAGCAGAGTGGACCACTCGAGTCCCTTCACTGAGAGGCTTTGTC 74357
QY 656 TTTTGCTCTTCAATTAATCTTGTCTGTCTATCTTTGTGTCACACTACCTTTATG 715
Db 74358 TTTTGCTCTTCAATTAATCTTGTCTGTCTATCTTTGTGTCACACTACCTTTATG 74417
QY 716 AGCTGTAACATCACTGAGAGGAGTGTGAGCTTCAATCTGAGAGTCA-ACAGACAGAA 774
Db 74418 AGCTGTAACATCACTGAGAGGAGTGTGAGCTTCAATCTGAGAGTCA-ACAGACAGAA 74477
QY 775 CCACTGGAAGGAACAAGAACTCCGATGTGCTGCTTTAAAGCTGTAACTCACTG 834
Db 74478 TCCACGAGAGGAACAAGAACTCCGATGTGCTGCTTTAAAGCTGTAACTCACTG 74537

```



QY 835 CGAAGCTGCAGCTTCACTCTCTGAAGTCAGTGCAGCAACCAACCCAGAGAGAGAA 894  
| | | | |  
Db 74538 TGAAGTCTGCGGCTTCACTCTCTGAAGTCAGGAGACCAAGAACCCAGAGAGAGAA 74597  
| | | | |  
QY 895 ACTCTGAGACACCTGATATATCTGAAAGAACAACTCCAGACACACCATCTTTGAGGCT 954  
| | | | |  
Db 74598 ACTCCAGACACATCTGAAATCTGAAAGAACAAACCCCTGAGACACACATCTTTAAGGCT 74657  
| | | | |  
QY 955 GTTACCTACACCGGAGGGCTGTGCTTCATCTTGAAGTCAGCAAGCAAGAACCA 1014  
| | | | |  
Db 74658 GTTACCTACCTGCAAGGGCTGTGCTTCATCTTGAAGTCAGTGCAGCAAGAACCA 74717  
| | | | |  
QY 1015 CCGAAGGAACAATTCAGACACAGTAGAATCTGATTTTGAATCT 1063  
| | | | |  
Db 74718 CCGAAGGAATTAATTCAGACACATTAATGATGATTAATTTAGGCT 74766  
| | | | |  
RESULT 10  
AC022034/c 195768 bp DNA linear PRI 01-JAN-2002  
LOCUS AC022034  
DEFINITION Homo sapiens chromosome 8, clone RP11-400K9, complete sequence.  
AC022034  
VERSION AC022034.7 GI:17426331  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 195768)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 8, clone RP11-400K9  
2 (bases 1 to 195768)  
Unpublished  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckert,J., Beda,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Deatellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,  
Macdonald,P., Margus,N., McKean,P., McGurk,A., McKernan,K.,  
McPheters,R., Meldrim,J., Menes,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,I.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rochnan,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vasilev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 195768)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Gilde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heatford,A., Horton,L., Hulme,W., Hulse,C., Jarocki,K.,  
Jones,C., Kann,L., Karatas,A., Kells,C., Lacroix,L.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Major,D., Margus,N., Matthews,C.,  
McCarthy,M., McKean,P., McKernan,K., McPheters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,N., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggillo,J., Vasilev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (01-JAN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 8, 2001 this sequence version replaced gi:17386392.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L2101  
Center clone name: 400\_K\_9  
-----  
FEATURES  
source  
Location/Qualifiers  
1. 195768  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-400K9"  
/clone\_1lb="RP11-400K9 Human Male BAC"  
914. 1047  
/rpt\_family="MIR3"  
1634. 1928  
/rpt\_family="AluSg"  
complement(5008..5192)  
/rpt\_family="MER5A"  
complement(5686..5778)  
/rpt\_family="MIR"  
complement(5817..5972)  
/rpt\_family="MIR3"  
6807..6894  
/rpt\_family="MIR3"  
7044. 7299  
/rpt\_family="MIR"  
complement(8103..8247)  
/rpt\_family="F1LM5"  
complement(9261..9553)  
/rpt\_family="AluSg"  
10329..10358  
/rpt\_family="AT\_rich"  
10794..10901  
/rpt\_family="MER96B"  
11864..11864  
/rpt\_family="AT\_rich"  
12585..12658  
/rpt\_family="MER113"  
complement(12696..12900)  
/rpt\_family="L1MC5"  
12901..13207  
/rpt\_family="AluSx"  
complement(13208..13252)  
/rpt\_family="L1MC5"  
13321..13718  
/rpt\_family="MER66A"  
complement(13719..14021)  
/rpt\_family="AluSx"  
14022..14091  
/rpt\_family="MER66A"  
complement(14192..14356)  
/rpt\_family="L1MC5"  
complement(14418..15549)  
/rpt\_family="L1MC5"  
complement(15545..15682)  
/rpt\_family="L1MD1"

repeat\_region complement(15665, .17136)  
/rpt\_family="L1MD2"  
repeat\_region 17137, .17440  
/rpt\_family="AluUb"  
repeat\_region complement(17441, 18005)  
/rpt\_family="L1MD2"  
repeat\_region complement(18002, .19359)  
/rpt\_family="L1MD2"  
repeat\_region 19384, .19509  
/rpt\_family="MIR"  
repeat\_region complement(19606, .19983)  
/rpt\_family="MER113"  
repeat\_region 19990, .20204  
/rpt\_family="L1MC4A"  
repeat\_region 20205, .20238  
/rpt\_family="CA)n"  
repeat\_region 20396, .20601  
/rpt\_family="L1MDa"  
repeat\_region complement(20602, .20753)  
/rpt\_family="TBE1C"  
repeat\_region complement(20756, 21978)  
/rpt\_family="MSTA-int"  
repeat\_region 21979, .22016  
/rpt\_family="TTTA)n"  
repeat\_region complement(22017, .22293)  
/rpt\_family="AluSx"  
repeat\_region complement(22294, 22709)  
/rpt\_family="MSTA-int"  
repeat\_region complement(22710, .23093)  
/rpt\_family="MSTA"  
repeat\_region 23150, .24986  
/rpt\_family="L1PA13"  
repeat\_region 24987, .25158  
/rpt\_family="AluUb"  
repeat\_region 25429, .25719  
/rpt\_family="AluSx"  
repeat\_region 26007, .26139  
/rpt\_family="L2"  
repeat\_region 26406, .26531  
/rpt\_family="L2"  
repeat\_region complement(26607, .26716)  
/rpt\_family="MER57A"  
repeat\_region complement(27622, .27755)  
/rpt\_family="MIR"  
repeat\_region 27775, .27877  
/rpt\_family="L2"  
repeat\_region complement(27929, .28003)  
/rpt\_family="MIR"  
repeat\_region complement(28066, .28275)  
/rpt\_family="MSTA"  
repeat\_region complement(28276, .31266)  
/rpt\_family="L1PA12"  
repeat\_region complement(31267, .31566)  
/rpt\_family="AluSg"  
repeat\_region complement(31567, .32933)  
/rpt\_family="L1PA12"  
repeat\_region complement(33237, .34054)  
/rpt\_family="L1PA12"  
repeat\_region complement(34153, .34631)  
/rpt\_family="L1PA12"  
repeat\_region complement(34637, .34770)  
/rpt\_family="FIAM\_C"  
repeat\_region complement(35125, .35415)  
/rpt\_family="AluSg"  
repeat\_region 35633, .35690  
/rpt\_family="MIR"  
repeat\_region 35713, .36004  
/rpt\_family="L2"  
repeat\_region complement(36568, .36845)  
/rpt\_family="L2"  
repeat\_region complement(36852, .37702)  
/rpt\_family="L1ME1"  
repeat\_region complement(37711, .38494)

Query Match 40.0%; Score 799.8; DB 8; Length 195768;  
Best Local Similarity 73.0%; Pred. No. 7, 4e-202;  
Matches 1233; Conservative 0; Mismatches 267; Indels 203; Gaps 11;  
QY 1 AAAGCTTGAAATCAGCAATGCTTCAACTCTTAACCACTCTGAGTGGC 60  
DB 124698 AAAGGCTTGAATGAGCAATGCTTCAACTCTTAACCACTCTGAGTGGC 124639  
QY 61 GACATGGCTTCCCTTTTCTAGGTCCTGTGACAGCCATCTGTATATGTGGCATTTGG 120  
DB 124638 AACATGGCTTCCCTTTTCTAGGTCCTGTGACAGCCATCTGTATATCTGCTTTGG 124579  
QY 121 GCCGTGATTTTAACTCTGTGCAAAATTTGTTCCCTAGAGTGAAGCCATCAAGCT 180  
DB 124578 GCCGTGATTTTAACTCTGTGCAAAATTTGTTCCCTAGAGTGAAGCCATCAAA-CT 124520  
QY 181 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCACTAACAACTTGTGAGAGACC 240  
DB 124519 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCACTAACAACTTGTACAGAGACC 124460  
QY 241 CCTGACCGACCCGCTGSCCTTTTCAATGSCCTTAAGCTCCCTGTGAGAGACCTAC 300  
DB 124459 CCTGACGTGACCTGCTGSCCTTTTCCCTGACCTAGAGAGCTCCCTGTGAGATCTAC 124400  
QY 301 CACTGAGGGGCCCTTTTCAACCCCTATCCAGAGAACTAGCTACAGGCTCATG-CC 359  
DB 124399 AACTGAGGGGCCCTTTTCAACCCCTATCCAGAGAACTAGCTACAGAGTCATTTGCC 124340  
QY 360 AATCCCAACAGACCTGGGCTGCTCTGTTTGAAGGGGATTTGAAGGTGAAGCAGCT 419  
DB 124339 AATTCACCAAGACATTTGGGGTGTCTGTTTGAAGGGGATTTGAAGGTGAAGCGCT 124280  
QY 420 GGGCTTCTGGGTCAGGTGGGACCTTGAGAACTTTTGTCTAGCTAAAGATTGTAAT 479  
DB 124279 GGGCTTCTGGGTCAGGAGGGGACTTGAGAACTTTTGTCTAGCTAAAGATTGTAAC 124220  
QY 480 GCACCAATCAGACCTGTGTCTAGCTAAAGATTGTAATGCAACCAATCAGCA----- 533  
DB 124219 ACACCAATCAGACCTGTGTCTAGCTAAAGATTGTAATGCAACCAATCAGCACTGT 124160  
QY 534 ----- 533  
DB 124159 AAAAAAGCAACATAGACCTGTGTCTAGCTAAAGATTGTAACGCAACCAATCAGCA 124100  
QY 534 -----CTGTGTAAT 544  
DB 124099 CTCTGTAATAATGACCAATCAGCACTTTGTGTAATGACCAATCAGTCTCTGTAAT 124040  
QY 545 GGACCAATCAGCAGATGTGGGCGGGGTCAATAAGGAGTAAACTGCGCACCCGAGC 604  
DB 124039 GGACCAATCAGCAGATGTGGGCGGGGTCAATAAGGAGTAAACTGCGCACCCGAGC 123980  
QY 605 CAGCAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTTTGTCT 664  
DB 123979 CAGCAGCAGCAACCTGCTTGGGTCCCTTCCACGCTGTGGAAGCTTTGTTCTTTTGTCT 123920  
QY 665 TCACATTAATCTTCTGCTGCTCACTTTGTGTGTCACACTTATATAGCTGTAAAC 724  
DB 123919 TCACATTAATCTTCTGCTGCTCACTTTTGTGTCACCTTTATATAGCTGTAAAC 123861  
QY 725 ACTCACTCGAGAGGTGTGTGCTTCACTTCTGTAAGTCAAC-AGACCAAGAACCACTGGA 783  
DB 123860 ACTCACTCGAGAGGTGTGTGCTTCACTTCTGTAAGTCAAGGAGACCAAGAACCCACAGG 123801  
QY 784 AGAACAAGAACTCCCATGTGTGCTTTAAGAGCTGTAACTCACTGAGAGCTCT 843  
DB 123800 AGAACAAGAACTCCCATGTGTGCTTTAAGAGCTGTAACTCACTGAGAGCTCT 123741  
QY 844 GCAGCTTCACTCTGAATCAGTGAACCAAACTCCACAGAGAGAACTTTGAC 903  
DB 123740 GTGGCTTCACTCTGAATCAGTGAACCAAACTCCACAGAGAGAACTTTGAC 123681

QY	904	ACACCTGTAATATCTGAAGAA	CAAACTCCAGACACACCATCTTTAGAGCTGTAACTCTC	963
Db	123680	ACATCTGAAACATCTGAAAGAA <th>CAAACTCCAGACACACCATCTTTAGAGCTGTAACTCTC</th> <td>123621</td>	CAAACTCCAGACACACCATCTTTAGAGCTGTAACTCTC	123621
QY	964	ACCCGAAAGGCTGTGGCTTCACT	TTCTTGAAGTCAACAGACCAGAACCCACCGAAGDA	1023
Db	123620	ACTGTGAAGGCTGTGGCTTCACT	TTCTTGAAGTCAACAGACCAGAACCCACCGAAGDA	123561
QY	1024	ACAAATCTCAGACACAGTAGA <th>AAATCTGTATTTTGTACTGTGGCTTCCAGGGTTACTCC</th> <td>1083</td>	AAATCTGTATTTTGTACTGTGGCTTCCAGGGTTACTCC	1083
Db	123560	ATCATATTACAGACACACTAG <th>CAAAAT-----</th> <td>123536</td>	CAAAAT-----	123536
QY	1084	AGTCATTGAAGTCTCCATTG <th>CACCTTTAAGAAAACAGAGATGCTTTGAGAGACATGT</th> <td>1143</td>	CACCTTTAAGAAAACAGAGATGCTTTGAGAGACATGT	1143
Db	123535	-----	TAATTAACCCAGAAAGTAAGTGTGGAACCTTTG	123501
QY	1144	TGGGAATTTGTTATGAGAC <th>CCAGGCTTGAGATGCACTATGAGGCATTTTCTGATCAAACTAGCT</th> <td>1203</td>	CCAGGCTTGAGATGCACTATGAGGCATTTTCTGATCAAACTAGCT	1203
Db	123500	ATTATATAGCTTTCAGTCA <th>GATGAGAGTATAGATGACAGAGCTTACTTGTGATTTGGTGTCTGA</th> <td>123441</td>	GATGAGAGTATAGATGACAGAGCTTACTTGTGATTTGGTGTCTGA	123441
QY	1204	GGAAACAGGCGCCAGGAAAT	ATATATATTAAGAAACAGATTTTTTGTAGACAGTAGTAGTCTT	1263
Db	123440	ATGGGGAG-----	AAAGTGTGGGAGCTGAAGTCTTAACTGTGGGAATCTGA	123394
QY	1264	TGCATCTGAGACATGTAG <th>ATTTATCAACGAAATTAATTAATTAAGAAAAATATAGCCAGGTCCAT</th> <td>1323</td>	ATTTATCAACGAAATTAATTAATTAAGAAAAATATAGCCAGGTCCAT	1323
Db	123393	TGCCATTTTCCAGTGTAG <th>TGTACAGAACAAATTAATTT-----ATTAAGCCCGGCGCAGT</th> <td>123341</td>	TGTACAGAACAAATTAATTT-----ATTAAGCCCGGCGCAGT	123341
QY	1324	GGCTCATGTGCTGTAAAT	CCACAGCCTTTGGGAGGCGCAAGGGGTGTGATATCAGAGGTCTAG	1383
Db	123340	AGCTCATGTGCTGTAAAT	CCACAGCCTTTGGGAGGCGCAAGGGGTGTGATATCAGAGGTCTAG	123281
QY	1384	CGTTCCAGACACAGCCTG <th>CGCAACATGATGAAAACCCGCTCTTACTTAATAAAAA--TACAAAAA</th> <td>1441</td>	CGCAACATGATGAAAACCCGCTCTTACTTAATAAAAA--TACAAAAA	1441
Db	123280	AGATTGAGACCAATCCCTG <th>GGCTTAACAGGTGAAACTGTCTCTAATAAAAAATACAAAAA</th> <td>123221</td>	GGCTTAACAGGTGAAACTGTCTCTAATAAAAAATACAAAAA	123221
QY	1442	TTAGCTGTGTGTGTG <th>TGTCGACATCTGTAATCCAG-TACTCAGAGGCTGTAGGCGGGG</th> <td>1500</td>	TGTCGACATCTGTAATCCAG-TACTCAGAGGCTGTAGGCGGGG	1500
Db	123220	TTAGCCAGGCGTGTGTG <th>TGTCGACATCTGTAATCCAG-TACTCAGAGGCTGTAGGCGGGG</th> <td>123161</td>	TGTCGACATCTGTAATCCAG-TACTCAGAGGCTGTAGGCGGGG	123161
QY	1501	AATCTCTTGAACCTTGG <th>AGGCGACAGAGTTGTGAGTGAAGCCAGATCACACCAACAGCAGCTCCG</th> <td>1560</td>	AGGCGACAGAGTTGTGAGTGAAGCCAGATCACACCAACAGCAGCTCCG	1560
Db	123160	AATGCGCTTGAACCCAG <th>AGGCGAGGCTTGTGAGTGAAGCCAGATCACACCAACAGCAGCTCCG</th> <td>123102</td>	AGGCGAGGCTTGTGAGTGAAGCCAGATCACACCAACAGCAGCTCCG	123102
QY	1561	TTCCTGGGTGACAGAG <th>CGAGACTGTCTCAAAAAAATTAATAATAATAATAATAATAATAATAA</th> <td>1620</td>	CGAGACTGTCTCAAAAAAATTAATAATAATAATAATAATAATAATAA	1620
Db	123101	GCTTGGGTGACAGAG <th>CGAGACTGTCTCTCAAAAAAATTAATAATAATAATAATAATAATAATAA</th> <td>123042</td>	CGAGACTGTCTCTCAAAAAAATTAATAATAATAATAATAATAATAATAA	123042
QY	1621	ATA 1623		
Db	123041	TTA 123039		
RESULT 11				
LOCUS	AC005613/c	62466 bp	DNA linear	PRI 04-SBP-1998
DEFINITION	Homo sapiens chromosome 5, BAC clone 114k9 (JBNL H94), complete			
ACCESSION	AC005613			
VERSION	AC005613.1	GI:3540152		
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.			
AUTHORS	1 (bases 1 to 62466) Kimmerly W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M., Kadner, K., Miguel, T., Miller, C., Pfluck, S., Pollard, M., Rojeski, H., Subramanian, S. and Martin, C. H.			
TITLE	Sequencing of human chromosome 5			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 62466)			

AUTHORS	Ricke, D. O.
TITLE	Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 62466)
AUTHORS	Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M., Kadner, K., Mignel, T., Miller, C., Plickuck, S., Pollard, M., Rojewski, H., Subramanian, S. and Martin, C.H.
TITLE	Direct Submission
JOURNAL	Submitted (04-SEP-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.
COMMENT	Sequence submitted by: DOE Joint Genome Institute.
FEATURES	Location/Qualifiers
source	1. .62466 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="5" /map="5q" /clone="114K3" /note="IBNL H94" 3810. .3936 /note="GRAIL 2 excellent exon, frame 2" complement(4364. .4417) /rpt_family="L1" 6262. .6639 /rpt_family="MLT2B2" complement(7533. .7788) /rpt_family="Alu" 10336. .10468 /rpt_family="MIR" 13065. .13197 /rpt_family="L1" 15009. .15297 /rpt_family="Alu" complement(11715. .17237) /note="GRAIL 2 excellent exon, frame 1" complement(11787. .17757) /rpt_family="MER3" complement(18497. .19099) /rpt_family="TTR12" complement(19289. .19376) /note="GRAIL 2 excellent exon, frame 2" complement(22605. .22883) /rpt_family="Alu" complement(24937. .25098) /note="GRAIL 2 excellent exon, frame 0" complement(26944. .27544) /rpt_family="TTR12" complement(28172. .28391) /rpt_family="MER43" 28588. .28762 /note="GRAIL 2 excellent exon, frame 0" 28866. .29046 /rpt_family="MER25" complement(30344. .30634) /rpt_family="Alu" 31121. .32400 /rpt_family="L1" 32804. .33080 /rpt_family="Alu" 34143. .34470 /rpt_family="L1" complement(37778. .41031) /rpt_family="L1" complement(38485. .39555) /note="GRAIL 2 excellent exon, frame 0" complement(41034. .41274) /rpt_family="Alu" 42715. .42951 /rpt_family="L1"

```

repeat_region      47785..47824
                    /note="(GAT)10"
                    /rpt_type=tandem
                    /rpt_unit="grat"
                    complement(47832..52964)
repeat_region      52127..52323
                    /rpt_family="L1"
                    complement(52323..52810)
repeat_region      52810..52999
                    /rpt_family="MER25"
                    complement(52999..53250)
repeat_region      53250..53678
                    /rpt_family="L1"
                    complement(53678..54006)
misc_feature        54006..54423
                    /note="Alu"
                    complement(54423..54468)
misc_feature        54468..56322
                    /note="GRAIL 2 excellent exon, frame 1"
                    56193..56322
repeat_region      56322..57743
                    /note="(GRAIL 2 excellent exon, frame 2"
                    57743..57782
                    /note="(AC)20"
                    /rpt_type=tandem
                    /rpt_unit="ac"
                    59472..59636
repeat_region      59636..60746
                    /rpt_family="MERS"
                    complement(60588..60746)
repeat_region      60746..61139
                    /rpt_family="L1"
                    61139..61455
repeat_region      61455..61455
                    /rpt_family="Alu"

```

## ORIGIN

Query Match 39.7%; Score 794.4; DB 8; Length 62466;  
 Best Local Similarity 86.0%; Pred. No. 2,1e-200;  
 Matches 937; Conservative 0; Mismatches 126; Indels 27; Gaps 4;

```

QY 2 AAGGCTTGAAGTACAGACAACTCTTCAACTCTTATACCAACCTGTGAGTGGCG 61
DB 19537 AAGGCTTGAAGTACAGACAACTCTTCAACTCTTATACCAACCTGTGAGTGGCG 19478
QY 62 ACATGCTTCTCCCTTTCTAGTCTGTGACAGCCATCTTGCTAATAGTCGATTGGG 121
DB 19477 ACATGCTTCTCCCTTTCTAGTCTGTGACAGCCATCTTGCTAATAGTCGATTGGG 19418
QY 122 CCGTATTTTAACTCTGTCAATTTGTTCCCTAGATGAGGCGCATCAAGCTA 181
DB 19417 CCGTATTTTAACTCTGTCAATTTGTTCCCTAGATGAGGCGCATCAAGCTA 19358
QY 182 CAGATGATCTTCAAAATGTAACCCCAATGAGCTCAACTAACAATCTGCTGAGACCC 241
DB 19357 CAGATGATCTTCAAAATGTAACCCCAATGAGCTCAACTAACAATCTGCTGAGACCC 19298
QY 242 CTGAGACGACCCGCTGCTTCAATGCTTAAAGCTCCCTCTGAGAGCACTA 301
DB 19297 CTGAGACGACCCGCTGCTTCAATGCTTAAAGCTCCCTCTGAGAGCACTA 19238
QY 302 ACTGAGGAGGCCCCCTTCTCAACCCCTATCCAGAGGAAGTATGACGCGCTATC-GCCA 360
DB 19237 ACTGAGGAGGCCCCCTTCTCAACCCCTATCCAGAGGAAGTATGACGCGCTATC 19178
QY 361 AATCCCAAGACAGCTGAGGCTGTCTGTTGAGAGGAGGAGTTGAGAGTGAACCCAGC 420
DB 19177 AATCCCAAGACAGCTGAGGCTGTCTGTTGAGAGGAGGAGTTGAGAGTGAACCCAGC 19118
QY 421 GCGTTCTGGGTCAAGTGGGAGCTTGAAGAACTTTGTGTCTAGCTTAAAGATTGTAAGT 480
DB 19117 GCGTTCTGGGTCAAGTGGGAGCTTGAAGAACTTTGTGTCTAGCTTAAAGATTGTAAGT 19058
QY 481 CACCAATCAGACACTGTGTCTAGCTTAAAGATTGTAAGTCAATCAGACACTGTG-- 538
DB 19057 CACCAATCAGACACTGTGTCTAGCTTAAAGATTGTAAGTCAATCAGACACTGTGAA 18998
QY 539 -----TAAATGAGCAATCAGCAGGATGTGGGCGGGGTCAAA 576
DB 18997 AATGAGCAATCAGTGTCTTAAATGAGCAATCAGTGGGCGGGGTCAAA 18938

```

```

QY 577 TAAGGAGTAAAGAACTGAGCCAGCCGAGCGAGTGGACCACTGAGGTCCTTCCA 636
DB 18937 TAAGGAGTAAAGAACTGAGCCAGCCGAGCGAGTGGACCACTGAGGTCCTTCCA 18878
QY 637 CACTGTGAAAGCTTTGTTCTTTTGTCTTCAATAAATCTGCTGCTCAATCTTTG 696
DB 18877 CACTGTGAAAGCTTTGTTCTTTTGTCTTCAATAAATCTGCTGCTCAATCTTTG 18818
QY 697 TGTCCACATCACTTTATAGCTGTAACTCACTGAGAGGTCGTGCTTCAATCTCTG 756
DB 18817 TGTCCACATCACTTTATAGCTGTAACTCACTGAGAGGTCGTGCTTCAATCTCTG 18758
QY 757 AAGTCAACAGACAGCAACCACTGAGAGGAAAGAAATCTCCGATGTCCTCTTAA 816
DB 18757 AAGTCAACAGACAGCAACCACTGAGAGGAAAGAAATCTCCGATGTCCTCTTAA 18699
QY 817 GAGCTGTAACTCACTGAGAGGTCGTGAGTCTGAGTCTCACTCTGAAGTC-AGTGAACACA 875
DB 18698 GAGCTGTAACTCACTGAGAGGTCGTGAGTCTGAGTCTCACTCTGAAGTCAGTGAACACA 18639
QY 876 AATCCCAAGAGAGAAAGAACTTGAACACACTGAATATCTGAAGAGCAAACTCCAGA 935
DB 18638 AATCCCAAGAGAGAGAAAGAACTTGAACACACTGAATATCTGAAGAGCAAACTCCAGA 18579
QY 936 CACACCATCTTGAAGCTGTAACTCACTCAACCGAAGGTCGTGCTTCAATCTTGAAGT 995
DB 18578 CACACCATCTTGAAGCTGTAACTCACTCAACCGAAGGTCGTGCTTCAATCTTGAAGT 18519
QY 996 CAGCAAGACCAAGAACCCAGCGAGAGAAATTCAGACAGTGAAGAAATCTGATT 1055
DB 18518 CAGCAAGACCAAGAACCCAGCGAGAGAAATTCAGACAGTGAAGAAATCTGATT 18459
QY 1056 TTTGATCTGT 1065
DB 18458 TTTGATCTGT 18449

```

RESULT 12  
 AC114962/c 107278 bp DNA linear PRI 05-NOV-2002  
 LOCUS Homo sapiens chromosome 5 clone RP11-206N2, complete sequence.  
 DEFINITION AC114962  
 ACCESSION AC114962  
 VERSION AC114962.2 GI:24580362  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 REFERENCE 1 (bases 1 to 107278)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 107278)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 REFERENCE 3 (bases 1 to 107278)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 On Nov 5, 2002 this sequence version replaced gi:19424423.  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www-shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.7% of Sequence;  
 Estimated Total Number of Errors is 0.2.  
 NOTE: This insert is not the entire sequence of the clone (entire  
 sequence is 201,6kb). It is clipped at the overlaps with AC027342

and AC010621. The number of bases overlapped with AC027342 is 5439  
and with AC010621 is 28528.

## FEATURES

source Location/Qualifiers

1..107278  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-206N2"

## ORIGIN

Query Match 39.7%; Score 793.6; DB 8; Length 107278;

Best Local Similarity 68.2%; Pred. No. 3.4e-200;

Matches 1339; Conservative 0; Mismatches 444; Indels 179; Gaps 10;

1 AAAAGCTTCTGAATACAGCAATGCTTTCACAACTTATACCACTCTGAGTTGGGC 60  
6414 AAAAGCTTCTGAATACAGCAATGCTTTCACAACTTATACCACTCTGAGTTGGGC 6355  
61 GACATGGCTTCCCTTCTAGTCTGTGACAGCATCTTGTATATAGTCATTTG 120  
6354 AACATGGCTTCCCTTCTAGTCTGTGACAGCATCTTGTATCTTACTCATCTTGG 6295  
121 GCCCTGATTTTAACTTCTGTGCAATTTGTTTCTCTAGATGAGGCAATCAAGCT 180  
6294 GCCCTGATTTTAACTTCTGTGCAATTTGTTTCTCTAGATGAGGCAATCAAGCC 6235  
181 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCACTAACAATCTTCTGAGAGAC 240  
6234 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCACTAACAATCTTCTGAGAGAC 6175  
241 CCGAGACCAACCGGCTGGCCCTTCAATGAGCTAAAGAGCTCCCTCTGAGAGCACTAC 300  
6174 CCGAGACCAACCGGCTGGCCCTTCAATGAGCTAAAGAGCTCCCTCTGAGAGCACTAC 6115  
301 CACTGAGGCCCCCTTCTTCAACCCCTATCCAGCAGAAAGTACAGCCGTCATCG -C 358  
6114 AACTGAGGCTCTCTTCTTTCCTTATCCAGCAGAAAGTACAGCCGTCATCGGCC 6055  
359 CAATCCCAACAGAGCTGGGGTGTCTGTTTGGAGGGGGATTTGAGAGTGAAGCCAGC 418  
6054 AAATCCCAACAGAGCTGGGGTGTCTGTTTGGAGGGGGATTTGAGAGTGAAGCCAGC 5995  
419 TGGGCTTCTGGGCTGAGGTGGGGACTTGGAGAACTTTGTCTAGCTAAAGATTGTA 478  
5994 TGGGCTTCTGGGCTGAGGTGGGGACTTGGAGAA-TTTTCTGTAGCTAAAGATTGTA 5936  
479 TGCACCAATCAGCACTCTGTCTAGCTAAAGATT----- 514  
5935 TGCACCAATCAGCACTCTGTCTAGCTAAAGATTGTAAACAACCAATCAGCACTCT 5876  
515 -GTAAATGACCAATCAGCACTCTGTAAATGACCAATCAGCAGATGTGGGGGGTGC 573  
5875 GTAAATGACCAATCAGCACTCTGTAAATGACCAATCAGCAGATGTGGGGGGTGC 5816  
574 AAATTAAGGGATTAAGAGCTGGCCACCGAGCAGACAGTGGCAACCACTGGGGTCCCTT 633  
5815 AAATTAAGGGATTAAGAGCTGGCCACCGAGCAGACAGTGGCAACCGCTCAGTCCCTT 5756  
634 CCAACAGTGGAAAGCTTGTCTTTTGTCTTTCACAAATTAATCTTGTCTGCTCATTTCT 693  
5755 CCAAGCTGTGGAAGCTTGTCTTTCATCTTTCGCAATTAATCTTGTCTGCTGCTCATCT 5696  
694 TTGTGTCAACATCTTTATGAGCTGTAAACATCACTGCGAGGGTCTGTGGCTTCATTTC 753  
5695 TTGTGTCCGACATTAATTAAGATGTGTCACTTACCGTGAAGGTCTGACGCTTCACTC 5636  
754 CTGAAGTCA----- 762  
5635 CTGAAGCCTAGAGCAATGAGCCACGAGAGGAAAAAGCACTCCAGAGTGCACCT 5576  
763 -----ACAGACC 769

5575 TTAAAGCTGTAACTACACCGCGAAGGTCTGACGCTTCACTCTTGAAGCCAGAGACC 5516  
770 ACGAACCCACTGGAAGAAAGAAAGAACTCCGAGTGTGCTTTTGAAGCTGTAACT 829  
5515 ACGAACCCACTGGAAGAAAGAAAGAAAGAACTCCGAGTGTGCTTTTGAAGCTGTAACT 5456  
830 CACTCGAAGCTCTGACGCTTCACTCTGAAAGTCACTGAGACCAAAACCCGCGAAGG 889  
5455 CACAGTGAAGGTCTGAAGCTTCACTCTGAAATGCGAGACCAAGAACCCGCGAAGG 5396  
890 AAGAACTCTGACACACCTGAAATATCTGAAGGAAAGAACTCCAGACCAACATCTTCA 949  
5395 AAGAACTCTGAAATATCTGAAATCTGAAGGAAAGAACTCCG-----ACATCTTTAA 5340  
950 GAGCTGTAACCTCAACCGCAAGGCTCTGTGCTTCACTTGAAGTCAAGCAAGCAAGA 1009  
5339 GAACTGTAACTCAACCGCAAGGCTCTGTGCTTCACTTGAAGTCAAGCAAGCAAGA 5280  
1010 ACCACCGGAGAAAGAAATTCAGACACAGTAC-----A 1045  
5279 ATCCACTGGAGAAATCAATTCGGAACATTAAGCCGCGCAAGTGCCTCAAGCTGCA 5220  
1046 AATCTGATTTTGAATCTGTGCTTCAAGGCTTCAAGTCAATGAAAGTCTCAATTGCA 1105  
5219 ATCCAGCACTTTTGGAGGTCAAGGAGTGAATCAAGGTTAGAGATCAAGACATTC 5160  
1106 GCCTTAAGGAAACAGAAATGTTTGGAGAGACATGTGGAAATTTTATGACCAAGCC 1165  
5159 CTGGCTAACACAGTAAACCCGCTCTTACTGAATAACAAATACTTGGCCGAGATGAG 5100  
1166 TTGAATGACATTAAGGCAATTTCTGATCAACCTTGAAGAGAGGCGCAGAAATATA 1225  
5099 TGGCACTCACTGTAGTCCCACTACTCCGAGAGGCTGAGAGAGGAAATGCTTAACTCG 5040  
1226 ATCTAAGGAAGACAGTTTGTAGACAGTAAAGTCTT-----GCATCTGAGACATGT 1279  
5039 GAGGCGGAGCTTGCAGTACGTAGATGTGCTGCTGATCTTCAAGCTTGGGTGAAGAG 4980  
1280 AGATTAACGCAATTAATTAAGAAATAATAGCCAGTGGCATGCTGATGCTGTAAT 1339  
4979 TGAATCTCATCTCAAAAAAATAATAGTGGGGTGAATGCTCAACCTGTAAAT 4920  
1340 CCCAGCACTTTGGAGGCCAAGGGGTGTGATCAAGAGTCAAGGCTTGAACCAAGCT 1399  
4919 CCCAGCACTTTGGAGGCCAAGGGGTGTGATCAAGAGTCAAGGCTTGAACCAAGCT 4860  
1400 GGCACAATGTGTAAACCCCGCTCTAATAA-----TACAAAAATTAGCTGGTGT 1453  
4859 GGCACAATGTGTAAACCCCGCTCTAATAAATAATACAAAAAATTAGCTGGTGT 4800  
1454 GGTGCAAGCATCTGTAAATCCAG-TACTCAGAGGCTGAGGAGGGAATCTTTGAAC 1512  
4799 GGGGTGAGGCTCTGTAGTCTGTGCTAATCCGAGGCTAAGGAGGAATGCGTGAAC 4740  
1513 TTGGAGGAGAGAGTTGCACTGAGCAAGATCAACCAAGCACTCCATCTTGGGTGACA 1572  
4739 CCGGAGGTGAGACTTGCATGAGCGAGATCAACCAAGCACTCCAGCTGGCGAACA 4680  
1573 GAGCGAGACTCTGTCTCAA-----AAAAAAGAAAGAAATATATATC 1626  
4679 GAGCGAGACTCTGTCTCAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 4620  
1627 AAGAAATTAAGAGTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1686  
4619 AAAAAATTAAGTGTAGGCTTAAAGCCCAACATATATATATTAATTAATTAATTAATG 4560  
1687 GTTTTACAGATTAATCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1746  
4559 GTTTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4500  
1747 TATCTCACTTATATAGTAAGAACTGAAGTACAGAAAGT 1788  
4499 TATATGTTCTTACAGAACTCAGATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 4458

RESULT 13	AC068682	184536 bp	DNA	linear	HTG 23-SEP-2000
LOCUS	AC068682	184536 bp	DNA	linear	HTG 23-SEP-2000
DEFINITION	Homo sapiens clone RP11-206N2, WORKING DRAFT SEQUENCE, 37 unordered pieces.				
ACCESSION	AC068682	3	GI:10280868		
VERSION	AC068682.3				
KEYWORDS	HTG, HTGS_PHASE1, HTGS_DRAFT.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia: Euteleostomi; Chordata: Vertebrata: Euteleostomi; Mammalia: Euteleostomi; Euteleostomi: Primates: Catarrhini; Homnidae, Homo.				
AUTHORS	1 (bases 1 to 184536)				
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Homo sapiens, clone RP11-206N2				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 184536)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepey,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Margis,N., McCarthy,M., McKean,P., McGuirk,A., McKernan,K., McPheters,R., Melidrum,T., Meneus,L., Milnova,T., Miranda,C., Mleaga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanì,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Steinge-Thomann,N., Stojanovic,N., Sudramanian,A., Talamas,J., Testaye,S., Theodore,J., Tyrrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zahouni,T., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Sep 23, 2000 this sequence version replaced gi:8247861. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
	----- Genome Center				
	Center: Whitehead Institute/ MIT Center for Genome Research				
	Center code: MIBR				
	Web site: http://www-seq.wi.mit.edu				
	Contact: sequence.submissions@genome.wi.mit.edu				
	----- Project Information				
	Center project name: U5802				
	Center clone name: 206_N_2				
	----- Summary Statistics				
	Sequencing vector: M13; M77815; 100% of reads				
	Chemistry: Dye-terminator Big Dye; 100% of reads				
	Assembly program: Phrap; version 0.960731				
	Consensus quality: 167659 bases at least Q40				
	Consensus quality: 175736 bases at least Q30				
	Consensus quality: 178914 bases at least Q20				
	Insert size: 15400; agarose-fp				
	Insert size: 180935; sum-of-coverage				
	Quality coverage: 5.0 in Q20 bases; agarose-fp				
	Quality coverage: 4.3 in Q20 bases; sum-of-coverage				
	-----				
	* NOTE: This is a 'working draft' sequence. It currently				
	* consists of 37 contigs. The true order of the pieces				
	* is not known and their order in this sequence record is				
	* arbitrary. Gaps between the contigs are represented as				
	* runs of N, but the exact sizes of the gaps are unknown.				

* This record will be updated with the finished sequence		
* as soon as it is available and the accession number will		
* be preserved.		
1	1154	contig of 1154 bp in length
1155	1254: gap of 100 bp	
1156	1255	2766: contig of 1512 bp in length
2767	2866: gap of 100 bp	
2867	4073: contig of 1207 bp in length	
4074	4173: gap of 100 bp	
4174	5191: contig of 1018 bp in length	
5192	5291: gap of 100 bp	
5292	7387: contig of 2096 bp in length	
7388	7487: gap of 100 bp	
7488	9087: contig of 1600 bp in length	
9088	9187: gap of 100 bp	
9188	10361: contig of 1174 bp in length	
10362	10461: gap of 100 bp	
10462	12065: contig of 1604 bp in length	
12066	12165: gap of 100 bp	
12166	13498: contig of 1333 bp in length	
13499	13598: gap of 100 bp	
13599	15085: contig of 1487 bp in length	
15086	15185: gap of 100 bp	
15186	17249: contig of 2064 bp in length	
17250	17349: gap of 100 bp	
17350	19243: contig of 1894 bp in length	
19244	19343: gap of 100 bp	
19344	21127: contig of 1784 bp in length	
21128	21227: gap of 100 bp	
21228	22345: contig of 2118 bp in length	
22346	22345: gap of 100 bp	
22446	25655: contig of 2520 bp in length	
25666	26065: gap of 100 bp	
26066	28195: contig of 2130 bp in length	
28196	28295: gap of 100 bp	
28296	32216: contig of 3921 bp in length	
32217	32316: gap of 100 bp	
32317	36554: contig of 4638 bp in length	
36955	37054: gap of 100 bp	
37055	40348: contig of 3294 bp in length	
40349	40448: gap of 100 bp	
40449	43556: contig of 3108 bp in length	
43557	43556: gap of 100 bp	
43557	47926: contig of 4270 bp in length	
47927	48026: gap of 100 bp	
48027	54079: contig of 6053 bp in length	
54080	54179: gap of 100 bp	
54180	58626: contig of 4783 bp in length	
58963	59621: gap of 100 bp	
59063	63739: contig of 4677 bp in length	
63740	63839: gap of 100 bp	
63840	71551: contig of 7712 bp in length	
71552	71651: gap of 100 bp	
71652	78218: contig of 6567 bp in length	
78219	78318: gap of 100 bp	
78319	84382: contig of 6664 bp in length	
84983	85082: gap of 100 bp	
85083	91930: contig of 6848 bp in length	
91931	92030: gap of 100 bp	
92031	98822: contig of 6892 bp in length	
98923	99022: gap of 100 bp	
99023	106496: contig of 7474 bp in length	
106497	106596: gap of 100 bp	
106597	114950: contig of 8354 bp in length	
114951	114950: gap of 100 bp	
115051	114950: contig of 9303 bp in length	
124354	124453: gap of 100 bp	
124354	132639: contig of 8186 bp in length	
132640	132739: gap of 100 bp	
132740	141572: contig of 8833 bp in length	
141573	141672: gap of 100 bp	
141673	151302: contig of 9630 bp in length	
151303	151402: gap of 100 bp	

```
FEATURES
    * 151403 161767: contig of 10365 bp in length
    * 161768 161867: gap of 100 bp
    * 161868 184536: contig of 22669 bp in length.
Location/Qualifiers
    source
        .184536
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /clone="RP11-206N2"
        /clone_11b="RPCT-11 Human Male BAC"
        1..1154
        /note="assembly_fragment"
        clone_end:SP6
        vector_side:left"
        1155..1254
        /estimated_length=100
        1255..2766
        /note="assembly_fragment"
        2767..2866
        /estimated_length=100
        2867..4073
        /note="assembly_fragment"
        4074..4173
        /estimated_length=100
        4174..5191
        /note="assembly_fragment"
        5192..5291
        /estimated_length=100
        5292..7387
        /note="assembly_fragment"
        7388..7487
        /estimated_length=100
        7488..9087
        /note="assembly_fragment"
        9088..9187
        /estimated_length=100
        9188..10361
        /note="assembly_fragment"
        10362..10461
        /estimated_length=100
        10462..12065
        /note="assembly_fragment"
        12066..12165
        /estimated_length=100
        12166..13498
        /note="assembly_fragment"
        13499..13598
        /estimated_length=100
        13599..15085
        /note="assembly_fragment"
        15086..15185
        /estimated_length=100
        15186..17249
        /note="assembly_fragment"
        17250..17349
        /estimated_length=100
        17350..19243
        /note="assembly_fragment"

Query Match      39.7%; Score 793.6; DB 14; Length 184536;
Best Local Similarity 68.2%; Pred. No. 3.3e-200;
Matches 1339; Conservative 0; Mismatches 444; Indels 179; Gaps 10;
```

```

Db 138446 GCCCTAATTTTAACTTTTGTGCAAAATTTTCTCTGGAATCAAGGCATCAAGCC 138505
Qy 181 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCAACTAACCTTGTGTGAGACC 240
Db 138506 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCAACTAACCTTGTGTGAGACC 138565
Qy 241 CCTGACCGACCCCGCTGCGCTTTTCAATGCGCTTAAAGACTCCCTCTGAGACACTAC 300
Db 138566 CTTGACCTGACTGCTGCGCTTTTCACTGCGCTTAAAGAGTTCCCTCTGAGACACTAC 138625
Qy 301 CACTGAGAGGCGCCCTTCTTCAACCCCTATCCAGAGAGTACTCAGCGCTCATG--C 358
Db 138626 AACTGAGAGGCTCTCTTCTTGTGCTTACAGAGAGTACTTAAAGCGTCTATGCGCC 138685
Qy 359 CAAATCCCAACAGACAGCTGGGGTGTCTCTGTTTGAAGGGGGAATTGAAGGTGAAGCCAC 418
Db 138686 AAATTCCAACAGCAAGTTGGGGCATCTCTTTTGAAGGGGGAATTGAAGGTGAAGCCAGC 138745
Qy 419 TGGGCTTCTGGGTCAGGTGGGGAATTGGAACTTTTGTCTAGCTTAAAGATTGTAA 478
Db 138746 TGGGCTTCTGGGTCAGGTGGGGAATTGGAA--TTTTCTGTCTAGCTTAAAGATTGTAA 138804
Qy 479 TGCACCAATCAGACACTGTGTCTAGCTTAAAGATT----- 514
Db 138805 TGCACCAATCAGACCTGTGTCTAGCTTAAAGATTGTTAAACACCAATCAGACACTCT 138864
Qy 515 -GTTAATGACCAATCAGACACTGTGTTAATGGAACCAATCAGAGAGTGGCGGGGTC 573
Db 138865 GTTAAATGACCAATCAGACACTGTGTTAATGGAACCAATCAGAGAGTGGCGGGGTC 138924
Qy 574 AAATTAAGGAGTAAATACTGGGCCACCGAGCCAGAGCTGGGCAACCCGCTCAGGTCCTCT 633
Db 138925 AAATTAAGGAGTAAATACTGGGCCACCGAGCCAGAGCTGGGCAACCCGCTCAGGTCCTCT 138984
Qy 634 CCACACTGTGGAGGCTTTGTCTTTGTCTTCAATTAATCTGTCTGCTCACTTCT 693
Db 138985 CCATGCTGTGGAAGCTTTGTCTTTCACTTGTGCAATTAATCTGTCTGCTCACTCT 139044
Qy 694 TTGTGTCCACACTCTTTATGAGCTGTAACTCACTCGAGAGGCTGTGGCTTCACTTC 753
Db 139045 TTGGGTCCGCACTAACTTATGATGTTGTCACTTACCGTGAAGGCTGAAGCTTCACTC 139104
Qy 754 CTGAAGTCA----- 762
Db 139105 CTGAAGCACTGAGACCATGAGCCACAGAGAGAGAGAGCACTCAGACGTGCCACT 139164
Qy 763 -----ACAGACC 769
Db 139165 TTAAAGCTGTAACTCTCACCGGAAGTCTGACCTTCACTCTTAAGCCAGCGAGACC 139224
Qy 770 ACAGACCACTGGAAGGAACAAAGAACTCCGATGTGCTGCTTTTAAAGCTGTAACT 829
Db 139225 ACAGACCACTGGAAGGAACAAAGAACTCTGACGACCACTTTTAAAGCTGTAACT 139284
Qy 830 CACTGCGAAGCTTGAAGCTTCACTCTTGAAGTCAGTGAGACCAAAACCACAGAGG 889
Db 139285 CACAGTGAAGGTCGGAAGCTTCACTCTGAAGTCAAGAGACCAAGAACCCACAGAGG 139344
Qy 890 AAGAACTCTGGAACACCTGAATATCTGAAGAGAAACAACTCCAGACACACCTTTTCA 949
Db 139345 AAGAACTCTGGAATATCTGAATATCTGAAGAGAAACAACTCCG-----ACATCTTTAA 139400
Qy 950 GAGCTGTAACTCACCGCAAGGCTGTGGCTTCACTTCTTGAAGTCAAGCAACCAAGA 1009
Db 139401 GAGCTGTAACTCACCGCAAGGCTGTGGCTTCTTGAAGTCAAGCAACCAAGA 139460
Qy 1010 ACCCAGCGAAGAAACAATTCAGACAGTAG-----A 1045
Db 139461 ATCCAGCGAAGAAACAATTCAGACAGTAG-----A 139520
Qy 1046 AATCTGATTTTGTATGTGTGCTTCCAGGTTATCTCAATGATCAAGTCTCATTTGCA 1105
Db 139521 ATCCAGCACTTGGGAGGTGAGGCGAGGTGATCAAGAGTTAAGGATCAAGCAATC 139580
```



```

QY 1106 GCCTTAGGAAACAGAGATGTTTGGAGGACATGTGGGATTTGTATGACCAAGC 1165
Db 139581 CTGGCTAAACACAGGAAACCCCGTCTTACTGAAATATCAAAAACATTACCGACATGG 139640
QY 1166 TTGAGATGCATATGGGCAATTTGTATCAAACTAGCTGAGAGAGGCGACGAAATATA 1225
Db 139641 TGGCAGTCACTGTAGTCCAGCTACTCGGAGGCTGAGGACGAGAGATCGTAAACCTG 139700
QY 1226 ATCTAAGGAAGACAGTTTGTATGACAGTAGTAGTCTTT-----GCATCTGAGACATGT 1279
Db 139701 GAGAGCGAGACTTGGACGTGAGCTGAGATTTGTCCACTGTACTCCAGCTTGAGTACGAG 139760
QY 1280 AGATTATCAAGCAATTATTAAGAAAAATATAGCAGATGCGATGGCTCATGCTGTAAT 1339
Db 139761 TGAGACTCCATCTCAAAAAAATATATATATATATATATATATATATATATATATAT 139820
QY 1340 CCCAGCACTTTGGAGGCGCAAGGGGTGTGATCAAGAGTCAAGCGCTTGAGACCAAGCT 1399
Db 139821 CCCAGCACTTTGGAGGCGCGAGGCGGCGACACCAAGTCAAGATCAAGACATCTCT 139880
QY 1400 GGCACATGTGTAACCCCGTCTCTACTAAAA-----TACAAAAATTAGCTGTGT 1453
Db 139881 GGCACATGTGTAACCCCGTCTCTACTAAAA-----TACAAAAATTAGCTGTGT 139940
QY 1454 GTTGGACAGCATCTGTATCCAG-TACTCAGAGAGGCTGAGGAGGGAATCTCTGTAAC 1512
Db 139941 GGGGGTGAAGCTCTGTATGCTCTGCTACTCCGAGGCTTAAAGGAGGAAATGGGCTGAAC 140000
QY 1513 TTGGAGGCGAGAGGTGCACTGAGCAGCAAGATCAACACAGCACTCCATCCGCTGAC 1572
Db 140001 CCGGAGGTGAGCTTGCATATGAGCGAGATCAACACAGCTCCAGCTCGGAGCA 140060
QY 1573 GAGCGAGCTCTCTCTCAAA-----AAAAAAGGAAAGGAAATATATATC 1626
Db 140061 GAGCGAGACTCCCTCTCAAAAGAAATATATATATATATATATATATATATATATAT 140120
QY 1627 AAGATATGAGAGGTGATCACTTATTCACACTTACTATGACACGAGCAATATACATAGT 1686
Db 140121 AAAAAATTAAGTGTAGGCTTAAAGCCAAACATATCAATATGTTACTTAAATGTAATG 140180
QY 1687 GTTTTATGATTAATCTATTAATCTTAAACATAGCCCTATGAAGTCAAGTCTGTAT 1746
Db 140181 GTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 140240
QY 1747 TATCTCACTTTATGATTAAGAACTGAATACAGAAAGT 1788
Db 140241 TATATGTTCTCTCAAGAAACTCAATCAATCAATCAATCAATCAATCAATCAATCA 140282

```

```

RESULT 14
LOCUS HSU116E9 25158 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone L10XNC01-116E9 on chromosome Xq22,
complete sequence.
ACCESSION 295333
VERSION 295333.2 GI:23304713
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Hominidae; Homo.
1 (bases 1 to 25158)
REFERENCE
AUTHORS Grafton-Cardi, D.
TITLE Direct Substitution
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonequest@sanger.ac.uk
On Sep 23, 2002 this sequence version replaced gi:2213445.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information

```

on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChX>  
 L10XNC01-116E9 is from the Lawrence Livermore National Laboratory flow-sorted  
 X chromosome cosmid library L10XNC01  
 VECTOR: law1816  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: vegas@sanger.ac.uk  
 -----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 Location/Qualifiers  
 1..25158  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="RZPD:L10XNC01E09116"  
 /db\_xref="taxon:9606"  
 /map="q22"  
 /clone="L10XNC01-116E9"  
 /clone\_1="L10XNC01"

ORIGIN

Query Match 39.2%; Score 785.2; DB 8; Length 25158;  
 Best Local Similarity 87.2%; Pred. No. 6e-198;  
 Matches 919; Conservative 0; Mismatches 128; Indels 7; Gaps 5;

```

QY 1 AAGGCTTCTGAATGACATGCTTCAAACTTTATACCACTCTGAGTTGGC 60
Db 9443 AAGGCTTCTGAATGACATGCTTCAAACTTTATACCACTCTGAGTTGGC 9502
QY 61 GACATGGCTTCCCTCTGAGTGTGAGCAGCATCTTATATGTCATTTGG 120
Db 9503 AATATGGCTTCCCTCTGAGTGTGAGCAGCATCTTATATGTCATTTGG 9562
QY 121 GCCGTATATTTTAACTCTTGTCAAAATTTGTTCTCTAGAGTGAAGCCATCAAGT 180
Db 9563 GCCTGTATTTTAACTCTTGTCAAAATTTGTTCTCTAGAGTGAAGCCATCAAGT 9622
QY 181 ACAGATGATCTTAACTTAATTAACCCCAATGAGCTCACTAACACTTGTGAGGACC 240
Db 9623 ACAGATGATCTTAACTTAATTAACCCCAATGAGCTCACTAACACTTGTGAGGACC 9682
QY 241 CCTGACGAGCCCGGCTGAGCTTCAAGTCCCTTAAGAGCTCCCTCTGAGGAGCACTAC 300
Db 9683 CCTGACGAGCCCGGCTGAGCTTCAAGTCCCTTAAGAGCTCCCTCTGAGGAGCACTAC 9742
QY 301 CACTGAGAGGCCCTTCTTCACTTATCAGACAGAAATAGTACAGAGGTCATC-GCC 359
Db 9743 AACTGAGAGGCCCTTCTTCACTTATCAGACAGAAATAGTACAGAGGTCATCAGC 9802
QY 360 AATTCACACAGCAGTGGGTGTCTCTGTTTGAAGGGGAGTTGAGAGGTGAAGCCAGCT 419
Db 9803 AATTCACACACAGTGGGTGTCTCTGTTTGAAGGGGAGTTGAGAGGTGAAGCCAGCT 9862
QY 420 GGGCTTCTGGGTGAGTGGGAGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 479
Db 9863 GGGCTTCTGGGTGAGTGGGAGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 9922
QY 480 GCACCAATCAGACTCTGTGTAGCTTAAAGATTGTAATGACCAATCAGACTCTGT 539
Db 9923 AACCAATCAGAGCTC--TATCTAGTAAAGTTGTAAATGACCAACAGCACTCTGT 9980

```

QY 540 AAAATGACCAATCAGCAGATGTGGCGGGTCAATAAGGAGTAAATAAGTGGCACC 599  
 Db 9981 AAAATGACCAATCAGCAGATGTGGCGGGTCAATAAGGAGTAAATAAGTGGCACC 10040  
 QY 600 CGAGCAGCAGTGGCAACCACTCGGGTCCCTTCACACTGTGGAAGCTTTGTTCTTTT 659  
 Db 10041 TGAAGCAGCAGCAGCAACCACTTTGGTCACTTCATGCTGTGGAAGCTTTGTTCTTTT 10100  
 QY 660 GCTCTTCACATTAATCTTGTGCTGCTCACTTTTGTGTCCACTACCTTTATGAGCT 719  
 Db 10101 GCTCTTCACATTAATCTTGTGCTGCTCACTTTTGTGTCCACTACCTTTATGAGCT 10160  
 QY 720 GTAACACTCACTCGAGGGTCTGTGCTTCATCTTCCGAAGTCA-ACAGCCACGAGCCCA 778  
 Db 10161 GTAACACTCACTCGAGGGTCTGTGCTTCATCTTCCGAAGTCA-ACAGCCACGAGCCCA 10220  
 QY 779 CTGGAAGGAACAAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACACTCACTCGAA 838  
 Db 10221 CTGAGGGAATTAACAACCTCCGATGTGCTCCCACTTTAAGAGCTGTAACACTCGAA 10280  
 QY 839 GCTCTGCACTTCACTCTCTGAAGTC-AGTGAACAACAACCAACCAAGAAAGAACT 897  
 Db 10281 AGTCTCGGGTTCACCTCTGAATCAAGAGATACCAATGAACCAAGAAAGAAAGAA-- 10338  
 QY 898 CTGGAACACCTGAATATCTGAAGGAACAACTCCAGACACCACTCTTCAGAGCTGTA 957  
 Db 10339 CTGGAACACCTGAATATCTGAAGGAATGAATCTTGACACCACTCTTTAAGAACTGTA 10398  
 QY 958 ACACTCAACGCAAGGGTCTGTGCTTCATCTTGAAGTCAAGACCAAGAAACCAACG 1017  
 Db 10399 ACACTCAACGCAAGGGTCTGTGCTTCATCTTGAAGTCAAGACCAAGAAACCAACG 10458  
 QY 1018 GAAGGAACCAATTCAGACACAGTGAAGAAATCTG 1051  
 Db 10459 GAAGGAACCAATTCAGACACAGTGAAGAAATGATG 10492

RESULT 15  
 AC092017 193041 bp DNA linear PRI 20-NOV-2001  
 LOCUS Homo sapiens chromosome 1 clone RP11-543B16, complete sequence.  
 AC092017 AL356972  
 AC092017.2 GI:17017614  
 HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 193041)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
 Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 193041)  
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and  
 Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-JUN-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 REFERENCE 3 (bases 1 to 193041)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
 Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-NOV-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 COMMENT On Nov 20, 2001 this sequence version replaced gi:14423605.  
 ----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UMG  
 Web site: <http://www.genome.washington.edu>  
 Contact: [umgchgs@u.washington.edu](mailto:umgchgs@u.washington.edu)

Drafting Center: SC  
 ----- Project Information  
 Center project name: chr-1  
 Center clone name: RP11-543B16 (sc0173)

----- Summary Statistics  
 Sequencing vector: plasmid; 108752; 100% of reads  
 Chemistry: Dye-terminator ET; 88% of reads  
 Chemistry: Dye-terminator Big Dye; 12% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 192648 bases at least Q40  
 Consensus quality: 193013 bases at least Q30  
 Consensus quality: 193041 bases at least Q20  
 Insert size: 200871; 9.7% error; agarose-fp  
 Insert size: 193041; sum-of-contigs  
 Quality coverage: 6.5x in Q20 bases; agarose-fp  
 Quality coverage: 6.7x in Q20 bases; sum-of-contigs

Overlapping Sequences:  
 5': Mapping in progress  
 3': Mapping in progress

----- Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BglII	SeqDerMap FngPrnt	SeqDerMap FngPrnt	SeqDerMap FngPrnt	HindIII
3118	3023	8696	8671	2515
2067	2109	6	<800	6382
9978	9634	62	<800	512
565	<800	3928	3905	449
2339	2376	12181	11850	4775
9773	9634	13497	13015	5390
1140	1152	1482	1462	485
1446	1435	1927	1935	7066
758	764	4640	4606	1525
				1475

1631	16312	8720	8671	775	771
6773	6774	1510	1462	887	891
376	<800	7700	7817	541	<800
3343	3344	3100	3152	6795	6975
8645	8582	723	730	5453	5382
2993	3023	1151	1071	2991	2987
5895	5784	15253	14879	1731	1769
1622	1606	7131	7188	4165	4095
5103	5075	2118	2064	3385	3214
6740	6774	1959	1935	664	<800
1905	1923	5491	5504	14342	14386
1637	1606	16	<800	8040	8002
1013	998	1090	1131	155	<800
526	<800	677	<800	1404	1338
11090	10902	57	<800	3025	3214
1175	1152	6052	6084	10582	10416
5066	5075	5140	5037	6178	6022
5214	5075	3074	3152	10309	10416
1475	1435	981	979	4191	4095
2313	2376	2669	2714	7361	7451
2340	2376	2223	2298	2667	2698
5769	5784	1309	1286	4804	4829
882	892	5371	5504	4832	4829
4356	4356	147	<800	3587	3548
6636	6774	15923	15755	3850	3829
4468	4356	1505	1462	2406	2428
3929	3914	8557	8671	9365	9375
7412	7695	3819	3905	734	771
15854	16312	1655	1642	11752	11688
6113	6108	6262	6354	310	<800
266	<800	23009	23575	5889	6022
102	<800	712	<800	102	<800
4345	4356	5827	6084	75	<800
603	<800	709	<800	1091	1079
1142	1152	1067	1071	4518	4481
6905	7396	2614	2714	4943	4829

2368	2376	6296	6311
3891	3914	4286	4333
34	<800	5345	5382
2135	2109	2905	2987
1421	1435		
FEATURES			
Location/Qualifiers			
source			
1..193041			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/chromosome="1"			
/clone_lib="RP11-543B16"			
/clone_lib="RP11-543B16"			
55137..55207			
misc_feature			
/note="Single subclone region"			
ORIGIN			
Query Match 39.1%; Score 782; DB 8; Length 193041;			
Best Local Similarity 86.9%; Pred. No. 4.1e-197;			
Matches 928; Conservative 0; Mismatches 110; Indels 30; Gaps 5;			
QY 1 AAAGGCTTGAATAGACAAATGCTTCAACTTTATACCACTCTGAGTTGGC 60			
DB 171360 AAAGGCTTGAATAGACAAATGCTTCAACTTTATACCACTCTGAGTTGGC 171419			
QY 61 GACATGGCTTCCCTTCTAGTCTGAGCAGCCATCTGTATATGTCATTGG 120			
DB 171420 AACATGGCTTCCCTTCTAGTCTGAGCAGCCATCTGTATATGTCATTGG 171479			
QY 121 GCCCTGATTTTAACTCTGTGTCAAATTTGTTCTCTAGATGAGCCATCAAGT 180			
DB 171480 GCCATATA-TTTTAACTCTTATCAATTTGTTCTCTAGATGAGCCATCAAGC 171538			
QY 181 ACAGATGATCTTAACTTAATGTAACCCCAATGAGCTCACTAACACTTCTGAGAGCC 240			
DB 171539 ACAGATGATCTTAACTTAATGTAACCCCAATGAGCTCACTAACACTTCTGAGAGCC 171598			
QY 241 CCTGACGAGCCCGGCTTTCATGAGCCCTAAAGAGCTCCCTGAGAGCACTAC 300			
DB 171599 CCTGACGAGCCCGGCTTTCATGAGCCCTAAAGAGCTCCCTGAGAGCACTAC 171658			
QY 301 CACTGAGAGCCCTTCTTCACTTATCCAGCAGAGTACTACAGCGTCA-TCGCC 359			
DB 171659 AACTGAGAGCCCTTCTTCACTTATCCAGCAGAGTACTACAGCGTCACTTGGCC 171718			
QY 360 AATCCCAACAGAGCTGGGGTGTCTGTTTGAAGGGGGAATTGAGAGTGAAGCCAGCT 419			
DB 171719 AATCCCAACAGAGCTGGGGTGTCTGTTTGAAGGGGGAATTGAGAGTGAAGCCAGCT 171776			
QY 420 GGGCTTGGGGTGGAGTGGGGAATTGAGAGCTTGTCTAGCTAAAGATTGAAT 479			
DB 171777 GGGCTTGGGGTGGAGTGGGGAATTGAGAGCTTGTCTAGCTAAAGATTGAAT 171836			
QY 480 GCACCAATGACACTGTGTCTAGCTAAAGATTGAATGATCACCATCAGCA----- 533			
DB 171837 ACACCAATGAGTGTGTCTAGCTAAAGATTGAATGATCACCATCAGCACTCTGT 171896			
QY 534 -----CTCTGTAATGAGCAATCAGAGATGAGCGGGGTCA 574			
DB 171897 AAAAAAGCAATTTGGCTCTCTGTAATGAGCAATCAGAGATGAGCGGGGTCA 171956			
QY 575 AATAAGGAGTAATAAATGAGCAAGCCAGAGTGGCAACCACTGGATCCCTTC 634			
DB 171957 AATAAGGAGTAATAAATGAGCAAGCCAGAGTGGCAACCACTGGATCCCTTC 172016			
QY 635 CACACTGTGAAGCTTGTCTTCTTCTTCAATTAATCTTCTGCTGCTCACTTCT 694			
DB 172017 CACACTGTGAAGCTTGTCTTCTTCTTCAATTAATCTTCTGCTGCTCACTTCT 172076			



```
misc_feature /estimated_length=100
              7435..9828
              /note="assembly_fragment"
gap          /estimated_length=100
              9829..9928
              /estimated_length=100
misc_feature /note="assembly_fragment"
              9929..14825
              /estimated_length=100
gap          /estimated_length=100
              14826..14925
              /estimated_length=100
misc_feature /estimated_length=100
              14926..21182
              /note="assembly_fragment"
gap          /estimated_length=100
              21183..21282
              /estimated_length=100
misc_feature /estimated_length=100
              21283..29127
              /note="assembly_fragment"
gap          /estimated_length=100
              29128..29227
              /estimated_length=100
misc_feature /estimated_length=100
              29228..38966
              /note="assembly_fragment"
gap          /estimated_length=100
              38967..39066
              /estimated_length=100
misc_feature /estimated_length=100
              39067..50624
              /note="assembly_fragment"
gap          /estimated_length=100
              50625..50724
              /estimated_length=100
misc_feature /estimated_length=100
              50725..67922
              /note="assembly_fragment"
gap          /estimated_length=100
              67923..68022
              /estimated_length=100
misc_feature /estimated_length=100
              68023..83588
              /note="assembly_fragment"
gap          /estimated_length=100
              83589..83688
              /estimated_length=100
misc_feature /estimated_length=100
              83689..98121
              /note="assembly_fragment"
gap          /estimated_length=100
              98122..98221
              /estimated_length=100
misc_feature /estimated_length=100
              98222..128190
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:left"
gap          128191..128290
              /estimated_length=100
misc_feature 128291..162397
              /note="assembly_fragment"
ORIGIN
```

```
Query Match      38.9%; Score 779; DB 14; Length 162397;
Best Local Similarity 86.9%; Pred. No. 2,6e-196;
Matches 917; Conservative 0; Mismatches 125; Indels 13; Gaps 5;
```

```
QY 1 AAAGGCTTGAATGACATGACATGCTTTGAACTTTATATACCACTCTGAGTGGGC 60
    |||||
DB 145936 AAAGTCTTGAATGACATGACATGCTTTGAACTTTATATACCACTCTGAGTGGGC 145877
    |||||
QY 61 GACATGGCTTCCCTTTTATAGTCTGTGACAGCCATTTGCTAATAGTGGCATTTG 120
    |||||
DB 145876 AACATGGCTTCCCTTTTATAGTCTGTGACAGCCATTTGCTAATAGTGGCATTTG 145817
    |||||
QY 121 GCCCTGATTTTAACTCTTGTGCAAAATTTGTTCTCTAGAGTGAAGCCATCAAGCT 180
    |||||
DB 145816 GCCCTGATTTTAACTCTTGTGCAAAATTTGTTCTCTAGAGTGAAGCCATCAAGCT 145757
    |||||
QY 181 ACAGATGATCTTAAATGTAACCCCAATGAGCTCACTAACACTTCTGTGAGGACC 240
    |||||
DB 145756 ACAGATGATCTTAAATGTAACCCCAATGAGCTCACTAACACTTCTGTGAGGACC 145697
    |||||
QY 241 CCTGAGCCAGCCGCTGGCCCTTTCAATGGCTTAAAGAGCTCCCTCTGAGAGCACTAC 300
    |||||
DB 145696 CCTGAGCCAGCCGCTGGCCCTTTCAATGGCTTAAAGAGCTCCCTCTGAGAGCACTAC 145637
    |||||
QY 301 CACTGAGGAGCCCTTCTTCAACCCCTATCAGAGGAGTATGACTAGCGGTGATGCGC- 359
    |||||
```

```
DB 145636 ATCTGAGGAGCCCTTCTTTGCCCC-ATCCAGAGGAAGTAGTAGACATGACCACT 145578
    |||||
QY 360 AAATCCCAACAGAGCTGGGAGTCTCTGTTTGAAGGGGGATTTGAGGTGAAGCCAGCT 419
    |||||
DB 145577 AATTCCCAACAGAGTGGGAGTCTCTGTTTGAAGGGGGATTTGAGGTGAAGCCAGCT 145518
    |||||
QY 420 GGGCTTCTGGGTCAGGTGGGGACTTTGAGAGACTTTTGTGCTAGCTTAAAGATTGTAAT 479
    |||||
DB 145517 GGGCTTCTGGGTCAGGTGGGGACTTTGAGAGACTTTTGTGCTAGCTTAAAGATTGTAAT 145458
    |||||
QY 480 GACCAATGACACTGTGTCTAGC-----TAAAGATTGTAATGACCAATCAG 531
    |||||
DB 145457 GCACCAATGACGCGCTCAATTAACACCAATCAGACTGTGTAATAACAGCAATCAG 145398
    |||||
QY 532 CACTGTGTAATAATGACCAATCAGAGATGTGGGGGGGTCAATTAAGGAGTAAATAAC 591
    |||||
DB 145397 CACTGTGTAATAATGACCAATCAGAGATGTGGGGGGGTCAATTAAGGAGTAAATAAC 145338
    |||||
QY 592 TGGCCACCCGAGCAGAGTGGCAACCACTCGGGTCCCTTTCACACTGTGAGAGCTTT 651
    |||||
DB 145337 TGGCCACCCGAGCAGAGTGGCAACCACTCGGGTCCCTTTCACACTGTGAGAGCTTT 145278
    |||||
QY 652 GTTCTTTTGTCTTCACAATTAATCTGTCTGCTCATTTCTTTGTCCACACTACCTT 711
    |||||
DB 145277 GTTCTTTTGTCTTCACAATTAATCTGTCTGCTCATTTCTTTGTCCACACTACCTT 145220
    |||||
QY 712 TATGAGCTGTAACTCACTGCGAGGGTCTGTGGCTTCACTTCTGAAATCAAC-AGAACA 770
    |||||
DB 145219 TATGAGCTGTAACTCACTGCGAGGGTCTGTGGCTTCACTTCTGAAATCAACAGCA 145160
    |||||
QY 771 CGAACCCACTGGAAGGAACAAAGAACTCCCGATGTGCTGCTTTAGAGCTGTAAACATC 830
    |||||
DB 145159 CAATCCCAACAGAGGAACCAACTCTGAGCGCCACCTTTAAAGCTGTAAATCTT 145100
    |||||
QY 831 ACTGCAAGACTGTGACCTTCACTCTGGAATGACGTGAGACCAACCAACCAAGGAAG 890
    |||||
DB 145099 ACTGCAAGACTGTGCGGGCTTCACTCTGGAATGAGCAATGAAACCAAGGAAG 145040
    |||||
QY 891 AGAACTTGGACACACTGTAATATTTGAAAGAACAACTCCAGACACACTTTTAC 950
    |||||
DB 145039 AGAACTTGGACACACTGTAATATTTGAAAGAACAACTCCAGACACACTTTTAC 144980
    |||||
QY 951 AGCTGTAACATCCAGCGCAAGGCTGTGGCTTCACTTTGAAAGTGAAGCAAGCAAG 1010
    |||||
DB 144979 AACTGTAACTCACTGCAAGGATCCGTGCTTCACTTTGAAAGTGAAGCAAGCAAG 144920
    |||||
QY 1011 CCCACCGAAGGAACAAATTTCCAGACACAGTAGA 1045
    |||||
DB 144919 CCCACCGAAGGAACCAATTTGAGACACTATGA 144885
    |||||
```

RESULT 17  
AC011328/c  
LOCUS Homo sapiens chromosome 10 clone RP11-295023, complete sequence.  
DEFINITION AC011328  
AC011328  
VERSION AC011328.11 GI:19744957  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 216031)  
AUTHORS Smith,D.R.  
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 216031)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (06-OCT-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

REFERENCE 3 (bases 1 to 216031)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUL-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

REFERENCE 4 (bases 1 to 216031)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

REFERENCE 5 (bases 1 to 216031)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

REFERENCE 6 (bases 1 to 216031)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

On Mar 27, 2002 this sequence version replaced gi:15021986.

----- Genome Center  
Center: Genome Therapeutics Corporation  
Center code: GTC  
Web site: <http://www.genomecorp.com/>  
Contact: gtc-seqcenter@genomecorp.com  
----- Project Information  
Center project name: hg031

----- Summary Statistics  
IMPORTANT: This sequence is the entire insert of clone RP11-295023. The true right end of clone RP11-506P9 is at 2000 in this sequence.

----- Sequencing Vector: N/A  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 990315

----- Location/Qualifiers  
source 1. 216031  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-295023"  
/clone\_id="RPC1-11"

ORIGIN

Query Match 38.9%; Score 779; DB 8; Length 216031;  
Best Local Similarity 86.9%; Pred. No. 2,66-196;  
Matches 917; Conservative 0; Mismatches 125; Indels 13; Gap 5;

QY 1 AAAGCTTCTGAATCAACAATGCTTTCAACTCTTTATCAACCTCTGAGTGGG 60  
187903 AAAGCTTCTGAATCAACAATGCTTTCAACTCTTTATCAACCTCTGAGTGGG 187844

QY 61 GACATGGCTTCCCTTTCTAGAGTCTCTGACAGCCATCTTGTCTAATAGCGATTGG 120  
187843 AACATGGCTTCCCTTTCTAGAGTCTCTGACAGCCATCTTGTCTAATAGCGATTGG 187784

QY 121 GCCGTATTTTAACTCTGATCAATTTGTTCTCTAGATCGAGGCATCAAGCT 180  
187783 GCCGTATTTTAACTCTGATCAATTTGTTCTCTAGATCGAGGCATCAAGCT 187724

QY 181 ACAGATGATCTTAAATGAATGAACCCCAATGAGCTCAACAACTTCTGAGAGCC 240  
187723 ACAGATGATCTTAAATGAATGAACCCCAATGAGCTCAACAACTTCTGAGAGCC 187664

QY 241 CTTGAGACGACCCCGTGGCCCTTTCAATGAGCTTAAGAGCTCCCTCTGAGAGACATAC 300  
187663 CTTGAGACGACCCCGTGGCCCTTTCAATGAGCTTAAGAGCTCCCTCTGAGAGACATAC 187604

QY 301 CACTGACAGGCGCTTTCTTCAACCCCTATCCAGACAGAGTAGTACAGCGGTCAATGCC 359

Db 187603 ATCTCAGGCGCCCTTCTTTGCCCC-ATTCAGACAGAAAGTACTAGAGAGTCAACACTT 187545

QY 360 AAATCCCAACAGACAGCTGGGGTGTCTCTGTTTGAAGGGGGATTGAGAGTGAAGCAGCT 419  
187544 AAATCCCAACAGACAGCTGGGGTGTCTCTGTTTGAAGGGGGATTGAGAGTGAAGCAGCT 187485

QY 420 GGGCTTCTGGGTCAAGTGGGGATTGAGAACTTTTGTCTTACTGTTAAAGATTGTAAT 479  
187484 GGGCTTCTGGGTCAAGTGGGGATTGAGAACTTTTGTCTTACTGTTAAAGATTGTAAT 187425

QY 480 GCACCAATAGACATCTGTGTCTAGC-----TAAAGATTGAATGACCAATCAG 531  
187424 GCACCAATAGACATCTGTGTCTAGC-----TAAAGATTGAATGACCAATCAG 187365

QY 532 CACTGTAAATATGAGACCAATGACAGAGATGTGGCGGGGTCAAAATGAGGATGAAAC 591  
187364 CACTGTAAATATGAGACCAATGACAGAGATGTGGCGGGGTCAAAATGAGGATGAAAC 187305

QY 592 TGGCCACCGGACGACAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGAAGCTTT 651  
187304 TGGCCACCGGACGACAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGAAGCTTT 187245

QY 652 GTTCTTTTCTCTTCAATAATTAATCTTGTCTGCTCAATTTGTGTCCACACTACTT 711  
187244 GTTCTTTTCTCTTCAATAATTAATCTTGTCTGCTCAATTTGTGTCCACACTACTT 187187

QY 712 TATGAGCTGTAACACTGACAGGGGTCTGTGCTTCAATTCCTGAAGTCAAC-AGACCA 770  
187186 TATGAGCTGTAACACTGACAGGGGTCTGTGCTTCAATTCCTGAAGTCAAC-AGACCA 187127

QY 771 CGAACCCTGTAAGAGGAACAAGAACTCCGATGTGCTCCTTTAAGAGCTGTACACTC 830  
187126 CGAACCCTGTAAGAGGAACAAGAACTCCGATGTGCTCCTTTAAGAGCTGTACACTC 187067

QY 831 ACTGCGAAGTCTGCAAGCTTCACTCCTGAAGTCAAGTGAACCAACCAAGGAAGGA 890  
187066 ACTGCGAAGTCTGCAAGCTTCACTCCTGAAGTCAAGTGAACCAACCAAGGAAGGA 187007

QY 891 AGAACCCTGTAAGAGGAACAAGAACTCCGATGTGCTCCTTTAAGAGCTGTACACTC 950  
187006 AGAACCCTGTAAGAGGAACAAGAACTCCGATGTGCTCCTTTAAGAGCTGTACACTC 186947

QY 951 AGCTGTAACTCACTCAGCGCAAGGCTGTGCTTCAATTTTGAAGTGAAGCAAGCAAGAA 1010  
186946 AGCTGTAACTCACTCAGCGCAAGGCTGTGCTTCAATTTTGAAGTGAAGCAAGCAAGAA 186887

QY 1011 CCCACCGGAAGGAACAATTCAGACACACTAGGA 1045  
186886 CCCACCGGAAGGAACAATTCAGACACACTAGGA 186852

RESULT 18  
AL390836/c 169482 bp DNA linear PRI 18-MAY-2005  
LOCUS Human DNA sequence from clone RP13-179C6 on chromosome X, complete  
DEFINITION sequence.  
ACCESSION AL390836  
VERSION AL390836  
KEYWORDS GI:11225794  
SOURCE HTG.  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
1 (bases 1 to 169482)  
REFERENCE Chapman,J.  
AUTHORS Direct Submission  
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
JOURNAL. Clone requests: [clone.requests@sanger.ac.uk](mailto:clone.requests@sanger.ac.uk)  
On Nov 19, 2000 this sequence version replaced gi:10277989.  
The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:  
 Em., EMBL; Sw., SWISSPROT; Tr., TRMBL; Wp., WORMBEP; Information  
 on the WORMBEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormep](http://www.sanger.ac.uk/Projects/C_elegans/wormep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping  
 Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX>  
 RP13-179C6 is from the library RP13-13.1 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 -----

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one subclone; and the assembly was confirmed by restriction digest,  
 except on the rare occasion of the clone being a YAC.

## FEATURES

source  
 1.169482  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="RP13-179C6"  
 /clone\_lib="RP13-13.1"

## ORIGIN

Query Match 38.9%; Score 778; DB 8; Length 169482;  
 Best Local Similarity 85.7%; Pred. No. 4.8e-196;  
 Matches 924; Conservative 0; Mismatches 120; Indels 34; Gaps 4;

QY 1 AAAAGCTTTGAAATGACCAATGCTTTTCAAACTTTATACCACTCTGAGATTGGG 60  
 DB 145912 AAAAGCTTTGAAATGACCAATGCTTTTCAAACTTTATACCACTCTGAGATTGGG 145853  
 QY 61 GACATGGCTTCCCTTTCTAGATCTGTGACAGCATCTTGTATATGTCGATTGG 120  
 DB 145852 AATCATGGCTTCTTCTTTCTAGATCTGTGACAGCATCTTGTATATGTCGATTGG 145793  
 QY 121 GCCCTGATTTTAACTCTTGTGTCAAATTTGTTCTCTAGATGAGCCATCAAGCT 180  
 DB 145792 GCCCTGATTTTAACTCTTGTGTCAAATTTGTTCTCTAGATGAGCCATCAAGCT 145733  
 QY 181 ACGATGATCTTCAAAATGTAACCCCAATGAGCTCACTAACCACTTCTGCTAGGA-C 239  
 DB 145732 ACGATGATCTTCAAAATGTAACCCCAATGAGCTCACTAACCACTTCTGCTAGGA-C 145673  
 QY 240 CCCTGACGAGCCCGCTGACCTTTCAATGGGCTAAAGAGCTCCCTCTGAGGACACTA 299  
 DB 145672 CCCTGACGAGCCCGCTGACCTTTCAATGGGCTAAAGAGCTCCCTCTGAGGACACTA 145613  
 QY 300 CCACTGACGAGCCCGCTTCTTCAACCCCTATCAGAGAGTACTACAGCGGTCAATCG-C 358  
 DB 145612 CCACTGACGAGCCCGCTTCTTCAACCCCTATCAGAGAGTACTACAGCGGTCAATCG-C 145553  
 QY 359 CAAATCCCAACGACGCTGGGCTGTCCTGTTGGAGGGGGATTGAGAGCTGAAGCCACG 418  
 DB 145552 CAAATCCCAACGACGCTGGGCTGTCCTGTTGGAGGGGGATTGAGAGCTGAAGCCACG 145493  
 QY 419 TGGGCTTCTGGGCTGAGGTGGGACTTGAGAACTTTTGTCTAGCTAAAGATTGTANA 478  
 DB 145492 TGGGCTTCTGGGCTGAGGTGGGACTTGAGAACTTTTGTCTAGCTAAAGATTGTANA 145433  
 QY 479 TGCACCAATCAGCACTC-----TGTGTCTAGCTA 507  
 DB 145432 TGCACCAATCAGCACTCCTCATTAATGACCAATCAGCACTCTGTAAATGACCAATCG 145373

QY 508 AAGATTGTAATGACCAATCAGCACTCTGTAATAATGACCAATCAGAGATGGGC 567  
 DB 145372 CACTCTGTAATAATGATCAATCAGCACTCTGTAATAATGACCAATCAGAGATGGGC 145313  
 QY 568 GGGGCTCAATTAAGGAGTAATAATGAGCCAGCCAGCAGAGTGGCAACCACTGGGT 627  
 DB 145312 AAGCCAAATTAAGGAGTAATAATGAGCCAGCCAGCAGAGTGGCAACCACTGGGT 145253  
 QY 628 CCCCTTCCACACTGTGGAGCTTTGTTCTTTTGTCTTCACAATAATCTTGTCTGCT 687  
 DB 145252 CCCCTTCCACACTGTGGAGCTTTGTTCTTTTGTCTTCACAATAATCTTGTCTGCT 145193  
 QY 688 CATTCTTTGTCACACTACCTTTATAGCTGTAACTCACTGCGAGGCTGTGGCT 747  
 DB 145192 TACTCTTTGGTTCACACTACCTTTATAGCTGTAACTCACTGCGAGGCTGTGGCT 145133  
 QY 748 TCATTCTGTAAGTCAAC-AGACCAAGAACCCACTGGAGGAACAAAGAACTCCGATGTG 806  
 DB 145132 TCATTCTGTAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 145073  
 QY 807 CTGCTTTAAGAGCTGTAACTCACTGCGAAGCTTGTGCACTTCTCTGAAGTCACT 866  
 DB 145072 CCACCTATTAAGAGCTGTAACTCACTGCGAAGCTTGTGCACTTCTCTGAAGTCACT 145013  
 QY 867 GAGACCAACCAACCCAGAGGAAGAACTGTGACACACCTGAAATATCTGAAGGACA 926  
 DB 145012 AAGACCAACCAACCCAGAGGAAGAACTGTGACACATCTGAACATCTGAAGGAATA 144953  
 QY 927 AACTCAGACACACCATCTTTTCAAGAGCTGTAACTCACTGCGAAGGCTGTGCTTCAT 986  
 DB 144952 AAGTCAAGACACACCATCTTTTCAAGAGCTGTAACTCACTGCGAAGGCTGTGCTTCAT 144893  
 QY 987 TCTTGAAGTCAAGACCAACCAACCCAGAGGAAGAACTGTGACACACCTGAAAGT 1044  
 DB 144892 TCTTGAAGTCAAGAGAGACCAACCAACCCAGAGGAAGAACTGTGACACACCTGAAAGT 144835

RESULT 19  
 HS785G19  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone RP4-785G19 on chromosome  
 20q11.1-11.23 contains the XPOPL gene for exportin tRNA (nuclear  
 export receptor for tRNAs) pseudogene 1, a UTRN (utrophin  
 (homologous to dystrophin)) pseudogene, the ASIP gene for agouti  
 (mouse)-signaling protein, the 3' end of AHcy gene for  
 S-adenosylhomocysteine hydrolase and three Cpg islands, complete  
 sequence.  
 AL035458  
 ACCESSION  
 AL035458.35 GI:6624641  
 VERSION  
 HTG; AHcy; ASIP; Cpg island; exportin; UTRN; utrophin; XPOPL.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.  
 1 (bases 1 to 115231)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Direct Submission  
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [vega@sanger.ac.uk](mailto:vega@sanger.ac.uk)  
 Clone request: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Dec 21, 1999 this sequence version replaced gi:6624642.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em., EMBL; Sw., SWISSPROT; Tr., TRMBL; Wp., WORMBEP; Information  
 on the WORMBEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormep](http://www.sanger.ac.uk/Projects/C_elegans/wormep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 20, constructed by the Sanger Centre Chromosome 20  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr20>  
 RP4-785G19 is from the library RP13-4 constructed by the group of



Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.

Location/Qualifiers

```
1. 115231
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCLP704G19785"
/db_xref="taxon:9606"
/chromosome="20"
/map="q11.1-11.23"
/clone="RP4-785G19"
/clone_11b="RPCL-4"
100
/feature="Clone_right_end: RP1-64K7"
complement(58003..60400)
/locus_tag="RP4-785G19.3-001"
/pseudo
/locus_tag="RP4-785G19.3-001"
/feature="match: proteins: O43592"
/pseudo
/codon_start=1
74336..74365
/feature="Tandem repeat. Single clone region. Assembly
confirmed by restriction digest data."
74892..74893
/feature="Tandem repeat"
join(complement(AL356299..16:10398..10517),
complement(94506..94646),complement(79351..79538))
/feature="AHCY"
/locus_tag="RP4-785G19.1-003"
join(complement(AL356299..16:10398..10517),
complement(94506..94646),complement(79351..79538))
/feature="AHCY"
/locus_tag="RP4-785G19.1-003"
/feature="match: ESTs: A1241007.1"
complement(79351)
/feature="AHCY"
/locus_tag="RP4-785G19.1-003"
complement(90461..91045)
/locus_tag="RP4-785G19.2-001"
/locus_tag="RP4-785G19.2-001"
/pseudo
complement(90461..91045)
/locus_tag="RP4-785G19.2-001"
/feature="match: proteins: AAH07860 O43583 O9CQJ6"
/pseudo
/codon_start=1
104873..113852
/feature="ASIP"
/locus_tag="RP4-785G19.4-001"
join(104873..105042,107340..107401,113499..113852)
/feature="ASIP"
/locus_tag="RP4-785G19.4-001"
/feature="match: ESTs: A1220203 A1378919 A1698587.1
AW206665.1 W02039 W32238
match: CDNAS: L06451 L06941 X99692 Y09877"
join(104883..105042,107340..107401,113499..113675)
/feature="ASIP"
/locus_tag="RP4-785G19.4-001"
```

misc\_feature

```
/standard_name="OTTHUMP0000030681"
/feature="match: proteins: P42127"
/codon_start=1
/protein_id="CAB96679.1"
/db_xref="GI:8953446"
/db_xref="Genbank:745"
/db_xref="GOA:P42127"
/db_xref="HSP:1HYK"
/db_xref="InterPro:IPR007733"
/db_xref="UniProt/Swiss-Pro: P42127"
/translation="MDYTRLLATLVLCFETNASHLPPEKLRDPSRSSVNL
LDPSVSIVALKSKSKQIGRAAKRSKSKKAAEMKYVRRRTPLSPCAVTRNSCKP
PAPACDPGCSQGRFPRASCSRVSLNC"
115231
/feature="Clone_right_end: RP4-785G19"
```

ORIGIN

```
Query Match 38.7%; Score 773.4; DB 8; Length 115231;
Best Local Similarity 84.0%; Pred. No. 8.1e-195;
Matches 952; Conservative 0; Mismatches 156; Indels 25; Gaps 6;

1 AAAGGCTCTGAAATCAGACCAATGCTTCAAACTCTATACCAACTCTGAGTTGGGC 60
1640 AAAGGTTCTGAAATCAGACCAATGCTTCAAACTCTATACCAACTCTGAGTTGGGC 1699
61 GACATGGCTTCCCTCTTCTAGGTCCTGTGACAGCCATCTTCTAATATGCAATTTGG 120
1700 AACGTGGCTTCCCTCTTCTAGGTCCTGTGACAGCCATCTTCTAATATGCAATTTGG 1759
121 GCCCTGATTTTAACTCTTCTGCAATTTGTTCTCTAGGATCGAGGCCATCAAGCT 180
1760 GCCCTGATTTTAACTCTTCTGCAATTTGTTCTCTAGGATCGAGGCCATCAAGCT 1819
181 AAGAGTATCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 240
1820 AAGAGTATCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 1879
241 CCTGACGACGCCGCTGCTGCTTCAATGAGCTTAAAGCTCCCTTGGAGACACTTAC 300
1880 CCTGACGACGCCGCTGCTGCTTCAATGAGCTTAAAGCTCCCTTGGAGACACTTAC 1939
301 CACTGACGACGCCGCTGCTGCTTCAATGAGCTTAAAGCTCCCTTGGAGACACTTAC -CC 359
1940 AACTGACGACGCCGCTGCTGCTTCAATGAGCTTAAAGCTCCCTTGGAGACACTTAC 1999
360 AATTCCTCAACAGAGTGGGGTGTCTGTTTGAAGGGGATTTGAGGTGAAGCCAGCT 419
2000 AATTCCTCAACAGAGTGGGGTGTCTGTTTGAAGGGGATTTGAGGTGAAGCCAGCT 2059
420 GGGCTT-CTGGGTCAAGTGGGGTGTCTGTTTGAAGGGGATTTGAGGTGAAGTTTAA 478
2060 GGGCTTCTGAGTCAAGTGGGGTGTCTGTTTGAAGGGGATTTGAGGTGAAGTTTAA 2119
479 TGCACCAATCAGAGTGTGTGTCTGTTTGAAGGGGATTTGAGGTGAAGTTTAA 538
2120 CGCACCAGTCAAGTGTGTGTCTGTTTGAAGGGGATTTGAGGTGAAGTTTAA 2162
539 TAAATATGACCAATCAGAGTGTGTGTCTGTTTGAAGGGGATTTGAGGTGAAGTTTAA 598
2163 TAAATATGACCAATCAGAGTGTGTGTCTGTTTGAAGGGGATTTGAGGTGAAGTTTAA 2222
599 CCGAGCAGAGTGGGACCACTCGGGTCCCTTCAACTGTGAGAGCTTTGTTCTTT 658
2223 CTGAGCAGATATGAGCAAGCCGTTCCGTTCCCTTCAACTGTGAGAGCTTTGTTCTTT 2282
659 TGGCTTCAATTAATCTTGTGTCTGTTTCAATCTTGTGTGTTTCAACTGTGAGAGCT 718
2283 CAGCTTCAATTAATCTTGTGTCTGTTTCAATCTTGTGTGTTTCAACTGTGAGAGCT 2342
719 TGTAACTCACTGAGAGGTTGTGTCTTCAATCTTGTGTGTTTCAACTGTGAGAGCT 777
2343 TGTAACTCACTCAAGAGGTTGTGTCTTCAATCTTGTGTGTTTCAACTGTGAGAGCT 2402
```

	*	11997	17768:	contig of 5772 bp in length
	*	17769	17868:	gap of 100 bp
	*	17869	33283:	contig of 15415 bp in length
	*	33284	33383:	gap of 100 bp
	*	33384	37004:	contig of 3621 bp in length
	*	37005	37104:	gap of 100 bp
	*	37105	72676:	contig of 35572 bp in length
	*	72677	72776:	gap of 100 bp
	*	72777	97002:	contig of 24226 bp in length
	*	97003	97102:	gap of 100 bp
	*	97103	121271:	contig of 24169 bp in length
	*	121272	121371:	gap of 100 bp
	*	121372	132660:	contig of 11289 bp in length
	*	132661	132760:	gap of 100 bp
	*	132761	145195:	contig of 12435 bp in length
	*	145196	145295:	gap of 100 bp
	*	145296	174784:	contig of 29489 bp in length
	*	174785	174884:	gap of 100 bp
	*	174885	182753:	contig of 7869 bp in length.
FEATURES				
source				
			1..182753	
			/organism="Homo sapiens"	
			/mol_type="Genomic DNA"	
			/db_xref="taxon:9606"	
			/chromosome="20"	
			/clone="RP11-263P6"	
			/clone_id="RPC1-11.1"	
			1..11896	
misc_feature			/note="assembly fragment:02743	
			fragment_chain:1	
			clone_end:77	
			vector_side:left	
			11997..17768	
misc_feature			/note="assembly fragment:00211	
			fragment_chain:1	
			17869..33283	
misc_feature			/note="assembly fragment:00124	
			fragment_chain:1	
			33384..37004	
misc_feature			/note="assembly fragment:02657	
			fragment_chain:1	
			37105..72676	
misc_feature			/note="assembly fragment:01648	
			fragment_chain:2	
			72777..97002	
misc_feature			/note="assembly fragment:02959	
			fragment_chain:2	
			97103..121271	
misc_feature			/note="assembly fragment:00003	
			fragment_chain:2	
			121372..132660	
misc_feature			/note="assembly fragment:02105	
			fragment_chain:3	
			132761..145195	
misc_feature			/note="assembly fragment:02149	
			fragment_chain:3	
			145296..174784	
misc_feature			/note="assembly fragment:02331	
			fragment_chain:3	
			clone_end:8P6	
			vector_side:right	
ORIGIN				
			38.7%: Score 773.4; DB 14; Length 182753;	
Query Match			Best Local Similarity 84.0%; Pred.No. 8e-195;	
Matches 952; Conservative			0; Mismatches 156; Indels 25; Gaps 6;	
OY				
1			AAAGGCTTCGAATCACAACATGGCTTCAACACTTTATACCAACCTCGAGTGAGGC	60
DB			AAGAAGTTTCGAATCACAACATGGCTTCAACACTTTATACCAACCTCGAGTGAGC	1821340

OY 61 GACATGGCTTCTCCCTTCTAGAGTCTGTGACAGCCATCTTGCTATATAGTGCATTTGG 120  
 DB 182133 AACGATGGCTTCTCCCTTCTAGAGTCTGTGACAGCCATCTTGCTATATAGTGCATTTGG 182074  
 OY 121 GCCCTGTATTTTAACTCTTGCTGCTAAATTTGTTCTCTAGAGTGCAGCCATCAAGCT 180  
 DB 182073 GCCCTGTATTTTAACTCTTGCTGCTAAATTTGTTCTCTAGAGTGCAGCCATCAAGCT 182014  
 OY 181 AAGATGTATTTTAACTCTTGCTGCTAAATTTGTTCTCTAGAGTGCAGCCATCAAGCT 240  
 DB 182013 AAGATGTATTTTAACTCTTGCTGCTAAATTTGTTCTCTAGAGTGCAGCCATCAAGCT 181954  
 OY 241 CCTGACGACCGCTGCGCTCTTCAATGAGCCCTTCAAGAGCTCCCTCTGAGAGCACTAC 300  
 DB 181953 CCTGACGACCGCTGCGCTCTTCAATGAGCCCTTCAAGAGCTCCCTCTGAGAGCACTAC 181894  
 OY 301 CACTGACGAGGCCCTTCTTCAACCCCTATCCAGAGAACTAGTACAGCGGTATCG-CC 359  
 DB 181893 CACTGACGAGGCCCTTCTTCAACCCCTATCCAGAGAACTAGTATGATGCTATGCGCC 181834  
 OY 360 AATATCCCAACAGAGCTGGGGTCTCTGTTTGAAGGGGGATTTGAGAGTGAAGCCAGCT 419  
 DB 181833 AATATCCCAACAGAGCTGGGGTCTCTGTTTGAAGGGGGATTTGAGAGTGAAGCCAGCT 181774  
 OY 420 GGGCTT-CTGGGTGAGGTGGGGAGCTTGAAGAACTTTTGTCTAGCTTAAAGATTTGTA 478  
 DB 181773 GGGCTT-CTGGGTGAGGTGGGGAGCTTGAAGAACTTTTGTCTAGCTTAAAGATTTGTA 181714  
 OY 479 TGCACCAATCAGCACTCTGTCTGCTAAGTAAAGTAAATGACCAATCAGCACTCTG 538  
 DB 181713 TGCACCAATCAGCACTCTGTCTGCTAAGTAAAGTAAATGACCAATCAGCACTCTG 181671  
 OY 539 TAAATGACCAATCAGCACTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598  
 DB 181670 TAAATGACCAATCAGCACTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181611  
 OY 599 CCGAGCAGAGGAGGAGCACTCGGGTCCCTTCCAGCACTGTGGAAGCTTTGTTCTT 658  
 DB 181610 CCGAGCAGAGGAGGAGCACTCGGGTCCCTTCCAGCACTGTGGAAGCTTTGTTCTT 181551  
 OY 659 TGGCTTCAATTAATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 718  
 DB 181550 TGGCTTCAATTAATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181491  
 OY 719 TGTAACTCACTGACGAGGCTCTGTGCTTCAATCTTGAAGTCAAC-AGACGACGAACTC 777  
 DB 181490 TGTAACTCACTGACGAGGCTCTGTGCTTCAATCTTGAAGTCAAC-AGACGACGAACTC 181431  
 OY 778 ACTGGAAGGAAGAAAGAACTCCGAGTGTGCTGCTTAAAGAGCTGAACACTGACGCTGA 837  
 DB 181430 ACTGGAAGGAAGAAAGAACTCCGAGTGTGCTGCTTAAAGAGCTGAACACTGACGCTGA 181371  
 OY 838 AGCTGACGCTTCACTCTGTAAGTCAAGTGAACCAACCAACCAACCAAGGAAGAACT 897  
 DB 181370 AGCTGACGCTTCACTCTGTAAGTCAAGTGAACCAACCAACCAACCAAGGAAGAACT 181311  
 OY 898 CTGGAACACCTGTAATCTGGAAGAAACAACTCAGACACACCATCTTTCAAGCTGTA 957  
 DB 181310 CTGGAACACCTGTAATCTGGAAGAAACAACTCAGACACACCATCTTTCAAGCTGTA 181255  
 OY 958 ACCTGACCGCAAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1016  
 DB 181254 ACCTGACCGCAAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181195  
 OY 1017 GGAAGGAACAAATTCAGACAGAGTGAAGAACTGTATTTTGTGCTGCTTCCAGG 1076  
 DB 181194 GGAAGGAACAAATTCAGACAGAGTGAAGAACTGTATTTTGTGCTGCTTCCAGG 181135  
 OY 1077 TTAATCCAGTCAATTAAGTCTCATTGACGCTTAAAGAAACAGAGAACTGTT 1129  
 DB 181134 TTAATCCAGTCAATTAAGTCTCATTGACGCTTAAAGAAACAGAGAACTGTT 181082

RESULT 21  
 AC084395  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 AC084395 174084 bp DNA linear HTG 30-OCT-2000  
 Homo sapiens chromosome 11 clone RP11-260L15, WORKING DRAFT  
 SEQUENCE, 19 unordered pieces.  
 AC084395.1 GI:11039612  
 HTG: HTGS PHASE1; HTGS\_DRAFT.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiinae; Homo.  
 1 (bases 1 to 174084)  
 Waterston, R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 174084)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (30-OCT-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H\_NH0260L15  
 ----- Summary Statistics -----  
 Sequencing vector: p13, 100%  
 Chemistry: Dye-terminator Big Dye, 0% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 16512 bases at least Q40  
 Consensus quality: 169389 bases at least Q30  
 Consensus quality: 170542 bases at least Q20  
 Insert size: 182000; agarose-fp  
 Insert size: 172284; sum-of-contigs  
 Quality coverage: 4.51 in Q20 bases; agarose-fp  
 Quality coverage: 4.79 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 19 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 1 7746: contig of 7746 bp in length  
 7747 7846: gap of unknown length  
 7847 16751: contig of 8905 bp in length  
 16752 16851: gap of unknown length  
 16852 23977: contig of 7126 bp in length  
 23978 24077: gap of unknown length  
 24078 34459: contig of 10332 bp in length  
 34460 34559: gap of unknown length  
 34560 44597: contig of 10038 bp in length  
 44598 44697: gap of unknown length  
 44698 55249: contig of 10552 bp in length  
 55250 55349: gap of unknown length  
 55350 66282: contig of 10932 bp in length  
 66282 66381: gap of unknown length  
 66382 80808: contig of 14427 bp in length  
 80809 91732: gap of unknown length  
 91733 91832: contig of 10824 bp in length  
 91833 106328: gap of unknown length  
 106329 106428: gap of unknown length

	*	106429	122673:	contig of 16245 bp in length
	*	122674	122773:	gap of unknown length
	*	122774	140593:	contig of 17820 bp in length
	*	140594	140693:	gap of unknown length
	*	140694	140692:	contig of 2399 bp in length
	*	141093	143192:	gap of unknown length
	*	141993	144628:	contig of 1436 bp in length
	*	144629	144728:	gap of unknown length
	*	144729	148312:	contig of 3584 bp in length
	*	148313	148412:	gap of unknown length
	*	148413	153703:	contig of 5291 bp in length
	*	153704	153803:	gap of unknown length
	*	153804	158969:	contig of 5166 bp in length
	*	158970	159069:	gap of unknown length
	*	159070	164878:	contig of 5809 bp in length
	*	164879	164978:	gap of unknown length
	*	164979	174084:	contig of 9106 bp in length.
FEATURES			Location/Qualifiers	
source			1..174084	
			/organism="homo sapiens"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:9606"	
			/chromosome="11"	
			/clone="RP11-260L15"	
misc_feature			1..7746	
			/note="assembly_name:Contig10"	
gap			7747..7846	
			/estimated_length=unknown	
misc_feature			7847..16751	
			/note="assembly_name:Contig11"	
gap			16752..16851	
			/estimated_length=unknown	
misc_feature			16852..23977	
			/note="assembly_name:Contig12"	
gap			23978..24077	
			/estimated_length=unknown	
misc_feature			24078..34459	
			/note="assembly_name:Contig13"	
gap			34460..34559	
			/estimated_length=unknown	
misc_feature			34560..44537	
			/note="assembly_name:Contig14"	
gap			44598..44697	
			/estimated_length=unknown	
misc_feature			44698..55249	
			/note="assembly_name:Contig15"	
gap			55250..55349	
			/estimated_length=unknown	
misc_feature			55350..66281	
			/note="assembly_name:Contig16"	
gap			66282..66381	
			/estimated_length=unknown	
misc_feature			66382..80808	
			/note="assembly_name:Contig17"	
gap			80809..80908	
			/estimated_length=unknown	
misc_feature			80909..91732	
			/note="assembly_name:Contig18"	
gap			91733..91832	
			/estimated_length=unknown	
misc_feature			91833..106328	
			/note="assembly_name:Contig19"	
gap			106329..106428	
			/estimated_length=unknown	
misc_feature			106429..122673	
			/note="assembly_name:Contig20"	
gap			122674..122773	
			/estimated_length=unknown	
misc_feature			122774..140593	
			/note="assembly_name:Contig21"	
gap			140594..140693	
			/estimated_length=unknown	
misc_feature			140694..143092	

[illegible]

```

Db      29990 AAAACGACCAATCAGCACTCTGTAAATGACCAATCATGAGTATGCGGAGGCGCA 30049
QY      575 AATAAGGAGTAAATCTGGCCACCGAGCCAGAGTGGCAACCACTGGGATCCCTTC 634
Db      30050 AATAAGGAGTAAATCTGGCCACCGAGCCAGAGTGGCAACCACTGGGATCCCTTC 30109
QY      635 CACACTGTGGAAGCTTTGTTCTTTTGTCTTCAATTAATCTGCTGCTCATCTT 694
Db      30110 CATGCTGTGGAAGCTTTGTTCTTTTGTCTTCAATTAATCTGCTGCTCATCTT 30169
QY      695 TGTGTCCACATCACTTTATGAGCTGTAACTACTCTGAGGAGTCTGTGCTTCATTC 754
Db      30170 TGGGTTCACACCACTTTATGAGCTGTAACTACTCTGAGGAGTCTGTGCTTCATTC 30229
QY      755 TGAAGTCAAC-AGACCAGCAACCCACTGTGAAGGAAGAAAGAACTCCGATGTGCTCCT 813
Db      30230 TGAAGTCAACGAGACCAACCAACCCAGGAGGAAGAAATGACTTGAACATGCACTT 30289
QY      814 TAAAGCTGTAACTACTCTGCAAGCTCTGCACTTCACTCTGAAAGTCACTGAGACCA 873
Db      30290 TAAAGCTGTAACTACTCTGCAAGCTCTGCACTTCACTCTGAAAGTCACTGAGACCA 30349
QY      874 CAATCCACCAAGGAAGAACTCTGCAACCACTGAAATATCTGAAGAACCAACTCCA 933
Db      30350 CAATCCACCAAGGAAGAAATTTGCAACATCTGAAATGAAACCAACTCTTG 30409
QY      934 GACACACCATCTTTTCAAGCTGTAACTCAACCGCAAGGCTGTGCTTCATCTTGA 993
Db      30410 GACACACCATCTTTTCAAGCTGTAACTCAACCGCAAGGCTGTGCTTCATCTTGA 30469
QY      994 GTGAGAGAGCAAGAACCCACCGAAGGAAGAAATTTCCAGACACAG 1040
Db      30470 GTGAGAGAGCAAGAACCCACCGAAGGAAGAACCAATTTCCGAGACACAG 30516

```

```

RESULT 22
AL160399 126595 bp DNA linear PRI 18-MAY-2005
LOCUS     Human DNA sequence from clone RP11-161N21 on chromosome 6, complete
DEFINITION
ACCESSION AL160399
VERSION   AL160399.13 GI:14018252
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homiidae; Homo.
           1 (bases 1 to 126595)
REFERENCE Holte, K.
TITLES    Direct Submission
JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
           Clone requests: clonerequest@sanger.ac.uk
           On May 11, 2001 this sequence version replaced gi:13751310.
           The following abbreviations are used to associate primary accession
           numbers given in the feature table with their source databases:
           EMT, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
           on the WORMPEP database can be found at
           http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
           was generated from part of bacterial clone configs of human
           chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
           Group. Further information can be found at
           http://www.sanger.ac.uk/HGP/Chr6
           RP11-161N21 is from the library RPI-11.1 constructed by the group
           of Pieter de Jong. For further details see
           http://www.chori.org/bacpac/home.htm
           VECTOR: pBACE3.6

```

```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

```

```

-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
Location/Qualifiers
1..126595
/molecule="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-161N21"
/clone_lib="RP11-161N21"
/clone_1ib="RP11-161N21"
100
/misc_feature
/note="Clone right_end: RP3-343124"
/misc_feature
/note="Sequence from clone PCR only. Sequence from
overlapping clone bA76119 (AL355854). Assembly confirmed
by restriction digest data. Sequence from uni-directional
dGTP big dye terminator reads only."
17753..17777
/misc_feature
/note="Sequence from clone PCR only. Sequence from
overlapping clone bA76119 (AL355854). Assembly confirmed
by restriction digest data. Sequence from uni-directional
dGTP big dye terminator reads only."
126496
/note="Clone left_end: RP3-493116"
ORIGIN
Query Match 38.6%; Score 771.8; DB 8; Length 126595;
Best Local Similarity 86.2%; Pred. No. 2,2e-194;
Matches 899; Conservative 0; Mismatches 117; Indels 27; Gaps 3;
QY 1 AAAAGCTTCTGAATATGACCAATGCTTTTCAAACTTTATACCACTCTGAGTTGGG 60
Db 56831 AAAAGCTTCTGAATATGACCAATGCTTTTCAAACTTTATACCACTCTGAGTTGGG 56890
QY 61 GACATGGCTTCCCTCTTCTAGAGTCTGAGCAGCATCTTGAATGTGCAATTTGG 120
Db 56891 AAGTGGCTTCCCTCTTCTAGAGTCTGAGCAGCATCTTGAATGTGCAATTTGG 56950
QY 121 GCGCTGATTTTAACTCTTGTGCAAAATTTGTTCTCTAGAGTCTGAGCAGCATCT 180
Db 56951 GCGCTGATTTTAACTCTTGTGCAAAATTTGTTCTCTAGAGTCTGAGCAGCATCT 57010
QY 181 ACAGATGATCTTCAATATGTAACCCCAATGAGCTCACTAACAATCTTGTGAGAGC 240
Db 57011 ACAGATGATCTTCAATATGTAACCCCAATGAGCTCACTAACAATCTTGTGAGAGC 57070
QY 241 CTTGAGACCAACCGGCTGCGCTTTCAATGAGGCTTAAGAGTCCCTCTGAGAGCACTAC 300
Db 57071 CTTGAGACCAACCGGCTGCGCTTTCAATGAGGCTTAAGAGTCCCTCTGAGAGCACTAC 57130
QY 301 CACTGAGAGGCGCTTCTTCAACCCCTATCAGAGAGGAAGTGAAGTGAAGTGAATC-GCC 359
Db 57131 AACTGAGAGGCGCTTCTTCAACCCCTATCAGAGAGGAAGTGAAGTGAAGTGAATC 57190
QY 360 AAATCCCAACAGAGAGTGGGAGTCTGTTTGAAGGGGAGTGAAGTGAAGTGAAGTGA 419
Db 57191 AAATCCCAACAGAGAGTGGGAGTCTGTTTGAAGGGGAGTGAAGTGAAGTGAAGTGA 57250
QY 420 GGGCTTCTGGGTGAGTGGGAGTCTGGAAGAACTTTGTGCTGAAGTGAAGTGAAGTGA 479
Db 57251 GGGCTTCTGGGTGAGTGGGAGTCTGGAAGAACTTTGTGCTGAAGTGAAGTGAAGTGA 57310
QY 480 GCAACATCAAGCACTGTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 522
Db 57311 ACACCAATCAAGTGTGTGTGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 57370
QY 523 -----ACCAATCAAGCACTGTGTAATAATGACCAATCAAGCAAGATGTGGGCGGCTCA 574

```

Db 57371 AAAAATAGACATTAAGCACTCTGTAAATGACCAATCAGCAGATGTGGTGGGCCCA 57430  
QY 575 AATTAAGGAGTAAATAATGGCCACCCGACGAGAGTGGCAACCATCGGGTCCCTTC 634  
Db 57431 AATTAAGGAGTAAATAATGGCCACCCGACGAGAGTGGCAACCATCGGGTCCCTTC 57490  
QY 635 CACACTGTGGAAGCTTGTCTTTTGTCTTCAATTAATCTTGCTGCTCATTTCTT 694  
Db 57491 CACGCTGTGAAGCTTTGTCTTTTGTCTTCAATTAATCTTGCTGCTCATTTCTT 57550  
QY 695 TGTGTCCACACTA CTTTATAGCTGTAACTCACTCAGAGGTCTGTGCTTCATTC 754  
Db 57551 TGGGCTGCACTACCTTATAGCTGTAACTCACTCAGAGGTCTGTGCTTCATTC 57610  
QY 755 TGAAGTCAAC-AAGACCGACGACCACTGGAAGGACAAAGACTCCGATGTGCTTC 813  
Db 57611 TGAAGGACGAGACCAACCAACCGACGAGAAATGACACTGTGACATGCACTT 57670  
QY 814 TAAGAGCTGTAACTCACTGCGAAGCTGTGACCTCACTCCGTAAGTGAAGACCA 873  
Db 57671 TAAGAGCTGTAACTCACTGCGAAGCTGTGACCTCACTCCGTAAGTGAAGACCA 57730  
QY 874 CAAACCCACGACGAGGAACTGTGACCACTGAAATATCTGAAGAACAACTCCA 933  
Db 57731 GGAACCAACGACGAGGAACTGTGACCACTGAAATATCTGAAGAACAACTCCA 57790  
QY 934 GACACACCATCTTTCAAGCTGTAACTCACTGCAAGGCTGTGCTTCATTTGAA 993  
Db 57791 GGACACCTATCTTTAACTGTAAGCTCAGCGTGAGGTGCATGCTTCATTTGAA 57850  
QY 994 GTGAGCAAGACCAAGAACCCACC 1016  
Db 57851 GTGACCGAGCTAAGAACCCACC 57873

RESULT 23  
AL355854 164532 bp DNA linear HTG 10-JUL-2001  
LOCUS Homo sapiens chromosome 6 clone RP11-76119, 11 unordered pieces.  
DEFINITION AL355854  
ACCESSION AL355854.8 GI:10186514  
VERSION HTG: HTGS PHASE1; HTGS\_CANCELLED.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1  
REFERENCE  
AUTHORS Burton, J.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone  
request: clonerequest@sanger.ac.uk  
On Sep 19, 2000 this sequence version replaced gi:9931763.  
COMMENT  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: BA76119  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 160164 bases at least Q40  
Consensus quality: 161830 bases at least Q30  
Consensus quality: 162686 bases at least Q20  
Insert size: 163532; sum-of-contigs  
Insert size: 167031; 6.6% error; agarose-fp  
Quality coverage: 4.89x in Q20 bases; sum-of-contigs Quality  
coverage: 5.05x in Q20 bases; agarose-fp

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
5323: contig of 5323 bp in length  
5324: gap of 100 bp  
5423: contig of 13021 bp in length  
5424: gap of 100 bp  
18445: contig of 3879 bp in length  
22423: contig of 32252 bp in length  
22524: gap of 100 bp  
54775: contig of 24890 bp in length  
54876: gap of 100 bp  
79765: contig of 21946 bp in length  
79766: gap of 100 bp  
101811: contig of 21946 bp in length  
101812: gap of 100 bp  
120185: contig of 18274 bp in length  
120186: gap of 100 bp  
120285: contig of 9402 bp in length  
129687: gap of 100 bp  
129788: contig of 3114 bp in length  
132901: gap of 100 bp  
132902: gap of 100 bp  
133001: contig of 21593 bp in length  
133002: contig of 21593 bp in length  
154594: gap of 100 bp  
154595: contig of 9838 bp in length.  
154695: contig of 9838 bp in length.

## FEATURES

## source

1.164532  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone\_lib="RP11-76119"  
/clone\_1ib="RP11-76119"  
1.5323  
/note="assembly fragment:01547  
fragment chain:1  
clone end:T7  
vector\_side:left"  
5424..18444  
/note="assembly fragment:00390  
fragment chain:1"  
18545..22423  
/note="assembly fragment:00760  
fragment chain:1"  
22524..54775  
/note="assembly fragment:00977  
fragment chain:1"  
54876..79765  
/note="assembly fragment:01205  
fragment chain:1"  
79866..101811  
/note="assembly fragment:02190  
fragment chain:2"  
101912..120185  
/note="assembly fragment:01164  
fragment chain:2"  
120286..129687  
/note="assembly fragment:00323"  
129788..132901  
/note="assembly fragment:00850"  
133002..154594  
/note="assembly fragment:01231  
fragment chain:3"  
154695..164532  
/note="assembly fragment:01221  
fragment chain:3  
clone end:SP6  
vector\_side:right"

## ORIGIN

Query Match 38.6%; Score 771.8; DB 14; Length 164532;  
 Best Local Similarity 86.2%; Pred. No. 2.2e-194;  
 Matches 899; Conservative 0; Mismatches 117; Indels 27; Gaps 3;

```

OY 1 AAAGGCTTGAATGAGCAATGCTTTCAAACTTTATACCACTCTGAGTTGGGC 60
    |||
DB 115346 AAAAGGCTTGAATGAGCAATGCTTTCAAACTTTATACCACTCTGAGTTGGGC 115405

OY 61 GACATGCTTCTCCCTTTCTAGGTCCTGTGACAGCCATCTTCTAATAGTGCATTGG 120
    |||
DB 115406 AACGTGGCTTCTCCCTTTCTAGGTCCTGTGACAGCCATCTTCTAATAGTGCATTGG 115465

OY 121 GCCCTGATTTTAACTCTGTGCTCAATTTGTTTCTCTAGGATGAGGCCATCAAGCT 180
    |||
DB 115466 GCACTATATTTTAACTCTGTGCTCAATTTGTTTCTCTAGGATGAGGCCATCAAGCT 115525

OY 181 ACAGATGATCTTACAATATGTAACCCCAATGAGCTCACTAACACTTCTGTGAGGACC 240
    |||
DB 115526 ACAGATGATCTTACAATATGTAACCCCAATGAGCTCACTAACACTTCTGTGAGGACC 115585

OY 241 CCTGACCCGACCGCTGCGCTTTCAATGCGCTTAAAGAGTCCCTCTGTGAGGACACTAC 300
    |||
DB 115586 CCTGACCCGACCCGCTGCGCTTTCAATGCGCTTAAAGAGTCCCTCTGTGAGGACACTAC 115645

OY 301 CACTGAGGGGCCCTTCTTCAACCCCTATCCAGAGGAGTACAGCGGTCATC-GCC 359
    |||
DB 115646 AACTGAGGGGCCCTTCTTCTTCCCTTATCTTGAAGAGTACAGGAGTGTATCGGCC 115705

OY 360 AAATCCCAACAGAGCTGGGGTCTCTGTTTGGAGGGGGATTGAGAGTGAAGCCAGCT 419
    |||
DB 115706 AAATCCCAACAGAGTGGGGTCTCTGTTTGGAGGGGGATTGAGAGTGAAGCGTGGCT 115765

OY 420 GGGCTTCTGGGTACAGTGGGGAATTGGAACAATTGTGTCTAGCTTAAAGATTGTAAT 479
    |||
DB 115766 GGGCTTCTGGGTACAGTGGGGAATTGGAACAATTGTGTCTAGCTTAAAGATTGTAAMC 115825

OY 480 GCACCAATGACGACTCTGTGTCTAGCTTAAAGATTGTAAMTGC----- 522
    |||
DB 115826 ACACCAATGACGACTCTGTGTCTAGCTTAAAGATTGTAAMTGCACCAATGACACTCTGT 115885

OY 523 -----ACCAATGACGACTCTGTAAATGACCAATGACAGAGTGGGGGTCTA 574
    |||
DB 115886 AAAATATGACCAATGACACTCTGTAAATGACCAATGACAGAGTGGGGGTCTA 115945

OY 575 AATTAAGGAGTAAATACTGCGCACCCGAGCCAGAGTGGCAACCACTCGGGTCCCTTC 634
    |||
DB 115946 AATTAAGGAGTAAATACTGCGCACCCGAGCCAGAGTGGCAACCACTCGGGGTCTCTTC 116005

OY 635 CACACTGTGGAAGTTGTTCTTTTGTCTTCAATAAATCTTGTGCTGCTCATCTT 694
    |||
DB 116006 CACACTGTGGAAGTTGTTCTTTTGTCTTCAATAAATCTTGTGCTGCTCATCTT 116065

OY 695 TGTGTCCACTACTCTTTATGAGCTGTAACTCACTGCGAGGGTCTGTGGCTTCATTCC 754
    |||
DB 116066 TGTGTCCACTACTCTTTATGAGCTGTAACTCACTGCGAAGGTCTGTGGCTTCCTCC 116125

OY 755 TGAAGTCAAC-AGAACAAGCCCACTGGAAGGAACAAGATCTCCGATGTGCTGCTT 813
    |||
DB 116126 TGAAGTCAAGCAAGCAACCACTGGAAGGAATGAACAATCTGTGACATGCACTT 116185

OY 814 TAAGAGCTGTAACTCACTGCGAAGCTTGTGCACTTCACTCCGTAAGTCAAGTGAACA 873
    |||
DB 116186 TAAGAGCTGTAACTCACTGCGAAGCTTGTGCACTTCACTCCGTAAGTCAAGTGAACA 116245

OY 874 CAATCCCAACGAGGAAGAACTCTGGAACAACCTGTAATATCTGAAGAAACAATCTCA 933
    |||
DB 116246 GGAATCCCAACGAGGAAGAACTCTGGAACAACCTGTAATATCTGAAGAAACAACCTCC 116305

OY 934 GACACCACTCTTCAAGAGCTGTAACTCACTGCGAAGGGTCTGTGGCTTCATTCTTGAA 993
    |||
DB 116306 GACACCACTCTTCAAGAGCTGTAACTCACTGCGAAGGGTCTGTGGCTTCATTCTTGAA 116365
  
```

OY 994 GTACGACAGACCAGAACCCACC 1016  
 |||  
 DB 116366 GTACGACAGACTTAAGAACCCACC 116388

RESULT 24  
 AP000924 172274 bp DNA linear PRI 17-JUN-2003  
 LOCUS Homo sapiens genomic DNA, chromosome 11 clone:RP11-758F15, complete  
 DEFINITION sequence.  
 ACCESSION AP000924  
 VERSION AP000924.7 GI:31790759  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Homo.

REFERENCE  
 1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Homo sapiens genomic DNA  
 Published Only in Database (1999)  
 2 (bases 1 to 172274)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Direct Submission  
 Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenho-chou,Tsukuba-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
 On Jun 16, 2003 this sequence version replaced gi:13359382.

COMMENT  
 Location/Qualifiers  
 source  
 1..172274  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11g"  
 /clone="RP11-758F15"

ORIGIN

Query Match 38.5%; Score 771; DB 8; Length 172274;  
 Best Local Similarity 86.3%; Pred. No. 3.5e-194;  
 Matches 921; Conservative 0; Mismatches 115; Indels 31; Gaps 5;

```

OY 1 AAAGGCTTGAATGAGCAATGCTTTCAAACTTTATACCACTCTGAGTTGGGC 60
    |||
DB 39185 AAAGGCTTCAAAATGAGCAATGCTTTCAAACTTTATACCACTCTGAGTTGGGC 39243

OY 61 GACATGCTTCTCCCTTTCTAGGTCCTGTGACAGCCATCTTAAATAGTGCATTGG 120
    |||
DB 39244 AACATGCTTCTCCCTTTCTAGGTCCTGTGACAGCCATCTTAAATAGTGCATTGG 39303

OY 121 GCCCTGATTTTAACTCTGTGCTCAATTTGTTTCTCTAGGATGAGGCCATCAAGCT 180
    |||
DB 39304 GCCCGATATTTTAACTCTGTGCTCAATTTGTTTCTCTAGGATGAGGCCATCAAGCT 39363

OY 181 ACAGATGATCTTACAATATGTAACCCCAATGAGCTCACTAACACTTCTGTGAGGACC 240
    |||
DB 39364 ACAGATGATCTTACAATATGTAACCCCAATGAGCTCACTAACACTTCTGTGAGGACC 39420

OY 241 CCTGACCCGACCGCTGCGCTTTCAATGCGCTTAAAGAGTCCCTCTGTGAGGACACTAC 300
    |||
DB 39421 CCTGACCCGACCGCTGCGCTTTCAATGCGCTTAAAGAGTCCCTCTGTGAGGACACTAC 39480

OY 301 CACTGAGGGGCCCTTCTTCAACCCCTATCCAGAGGAGTACAGCGGTCAATC-GCC 359
    |||
DB 39481 AACTGAGGGGTCTTCTTTTGTCCCTATCTCAAGAGGAGTACAGCGGTCAATCACC 39540

OY 360 AAATCCCAACGAGAGTGGGGTGTCTGTTTGGAGGGGGATTGAAGAGTGAAGCCAGCT 419
    |||
  
```



|||||  
Db 39541 AATTCACACAGAGTGGGGTCTCTTGAAGTGGGAGATTGAGAGCCAGT 39600  
420 GGGCTCTGGGTGAGGGGAGCTTGAGAACTTTGTCTCTAAGATTGTAAT 479  
Db 39601 GGGCTCTGGGTGAGGGGAGCTTGAGAACTTTGTCTCTAAGATTGTAAT 39660  
Qy 480 GCACCAATGACACTCTGTCTAGCTAAGATTTGAATGC----- 522  
Db 39661 ACACCAATGACACTCTGTCTAGCTAAGATTTGAATGCACCAATGACACTCTGT 39720  
Qy 523 -----ACCAATGACACTCTGTAAATGACCAATGACAGATGTGGGCGGTCA 574  
Db 39721 AAAACGACCAATGACACTCTGTAAATGACCAATGACAGATGTGGGCGGTCA 39780  
Qy 575 AATAAGGAGTAAATAATGCGCCACCGGAGCAGAGTGGCAACCCACTGGGTCCTTC 634  
Db 39781 AATAAGGAGTAAATAATGCGCCACCGGAGTGGCAACCCACTGGGTCCTTC 39840  
Qy 635 CACACTGTGAGACTTGTCTTTGCTCTTCAATTAATCTTGCTGTCTCACTTCT 694  
Db 39841 CAGCTGTGAGACTTGTCTTTGCTCTTCAATTAATCTTGCTGTCTCACTTCT 39900  
Qy 695 TGTGTCAACACTCTTTATGAGCTGTAACTCACTGCGAGGCTCTGTCTTCACTTC 754  
Db 39901 TGGGTCAACCACTTTATGAGCTGTAACTCACTGCGAGGCTCTTCACTTC 39960  
Qy 755 TGAAGTCAAC-AGACCAACCACTGAGAGCAAAAGAACTCCCGATGTCTGCTT 813  
Db 39961 TGAAGTCAAGAGCAACCACTGAGAGCAAAAGAACTCCCGATGTCTGCTT 40020  
Qy 814 TGAAGTCAAGAGCACTGAGAGCTGAGAGCTTCACTCTGAGTCAAGTCAAGCA 873  
Db 40021 TGAAGTCAAGAGCACTGAGAGCTGAGAGCTTCACTCTGAGTCAAGCA 40080  
Qy 874 CAAACCCACAGAGAGAACTCTGACACACTGAAATCTGAAGAGCAAACTCTCA 933  
Db 40081 CGAACCACTGAGAGAGAAATTTGGACACATCTGAATCTGAATGAACAACTCTG 40140  
Qy 934 GACACCACTTTTCAAGCTGTAACTCACTGAGAGGCTCTGTCTTCACTTGA 993  
Db 40141 GACACCACTTTTCAAGCTGTAACTCACTGAGAGGCTCTGTCTTCACTTGA 40200  
Qy 994 GTACAGAGAGCAAGAACCCACCGAGAGAAATTTCCAGACACAG 1040  
Db 40201 GTACAGAGAGCAAGAACCCACCGAGAGAAATTTCCAGACACAG 40247

RESULT 25  
AP002788 174311 bp DNA linear PRI 15-MAR-2003  
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-347E10,  
DEFINITION complete sequence.  
ACCESSION AP002788  
VERSION AP002788.3 GI:13429927  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE  
1 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Tocoli, Y., Watanabe, H. and Sakaki, Y.  
Homo sapiens genomic DNA  
Published Only in Database (2000)  
2 (bases 1 to 174311)  
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Tocoli, Y., Watanabe, H. and Sakaki, Y.  
Direct Submision  
Submitted (11-JUL-2000) Masahita Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: hattori@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
On Mar 21, 2001 this sequence version replaced gi:9845043.  
COMMENT  
FEATURES  
Source  
1. 174311  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11q"  
/clone="RP11-347E10"

ORIGIN  
Query Match 38.5%; Score 771; DB 8; Length 174311;  
Best Local Similarity 86.3%; Pred. No. 3.5e-194;  
Matches 921; Conservative 0; Mismatches 115; Indels 31; Gaps 5;

1 AAGGCTTGTGAATGAGCAATGCTTTCAAACTCTTATACCACTCTGAGTTGGC 60  
147400 AAGGCTTCCAAATATGACAAATCTTTCAACTCTTATACCACTCTGAGTT- GGC 147458  
Qy 61 GACATGGCTTCTCCCTTTCTAGGCTCTGTGACAGCCATCTTGTCTAATAGTGCATTTGG 120  
Db 147459 AACATGGCTTCTCCCTTTCTAGGCTCTGTGACAGCCATCTTGTCTAATAGTGCATTTGG 147518  
Qy 121 GCCCTGATTTTAACTCTGTGTCAATTTGTTCTCTAGAGTGAAGGCCATCAAGCT 180  
Db 147519 GCCCTGATTTTAACTCTGTGTCAATTTGTTCTCTAGAGTGAAGGCCATCAAGCT 147578  
Qy 181 ACAGATGATCTTCAATGATGAACCCCAATGAGCTCACTAACAATCTCTGAGAGACC 240  
Db 147579 ACAGATGATCTTCAATGATGAACCCCAATGAGCTCACTAACAATCTCTGAGAGACC 147635  
Qy 241 CTTGACCGAACCGCTGCTCTTCAATGAGCTTCAATGAGCTTCTGAGAGCACTAC 300  
Db 147636 CTTGACCGAACCGCTGCTCTTCAATGAGCTTCAATGAGCTTCTGAGAGCACTAC 147695  
Qy 301 CACTGAGAGGCCCTTCTTCAACCCCTATCCAGAGAGATGATCAAGGCTATC- GCC 359  
Db 147696 AACTGAGAGGCTCTTCTTCTTCTGCTTCAAGAGAGATGATCAAGGCTATC- GCC 147755  
Qy 360 AATCCCAACAGAGCTGAGGAGTCTCTGTTTGGAGGGGAGATTGAGAGTGAAGCCAGCT 419  
Db 147756 AATCCCAACAGAGCTGAGGAGTCTCTGTTTGGAGGGGAGATTGAGAGTGAAGCCAGCT 147815  
Qy 420 GGGCTTCTGGGTCAAGGTGGGAACTTGAAGAACTTTGTCTAAGTAAAGATTGTAAT 479  
Db 147816 GGGCTTCTGGGTCAAGGTGGGAACTTGAAGAACTTTGTCTAAGTAAAGATTGTAAT 147875  
Qy 480 GCACCAATGACACTCTGTCTAGCTAAGATTTGAATGC----- 522  
Db 147876 ACACCAATGACACTCTGTCTAGCTAAGATTTGAATGCACCAATGACACTCTGT 147935  
Qy 523 -----ACCAATGACACTCTGTAAATGACCAATGACAGATGTGGGCGGTCA 574  
Db 147936 AAAACGACCAATGACACTCTGTAAATGACCAATGACAGATGTGGGCGGTCA 147995  
Qy 575 AATAAGGAGTAAATAATGCGCCACCGGAGCAGAGTGGCAACCCACTGGGTCCTTC 634  
Db 147996 AATAAGGAGTAAATAATGCGCCACCGGAGCAGAGTGGCAACCCACTGGGTCCTTC 148055  
Qy 635 CACACTGTGAGACTTGTCTTTGCTCTTCAATTAATCTTGCTGTCTCACTTCT 694  
Db 148056 CAGCTGTGAGACTTGTCTTTGCTCTTCAATTAATCTTGCTGTCTCACTTCT 148115  
Qy 695 TGTGTCAACACTCTTTATGAGCTGTAACTCACTGCGAGGCTCTGTCTTCACTTC 754  
Db 148116 TGTGTCAACACTCTTTATGAGCTGTAACTCACTGCGAGGCTCTGTCTTCACTTC 148175  
Qy 755 TGAAGTCAAC-AGACCAACCACTGAGAGCAAAAGAACTCCCGATGTCTGCTT 813  
Db 148176 TGAAGTCAAGAGCAACCACTGAGAGCAAAAGAACTCCCGATGTCTGCTT 148235

OY		814	TAAGAGCTGTAACTCACTCATGGGAAGCTTCGACGCTTCACTCCTGAAAGTCAAGAACCA	873
Db		148236	TAAAGCTGTAACTCACTCATGGGAAGCTTCGACGCTTCACTCCTGAAAGTCAAGAACCA	1482355
OY		874	CAAAACCCACCAAGAAGAAACTGTGCACACCTGAATATCTGAAGAAACAACTCCA	933
Db		148236	CGAACCACCTGGAAGGAATAATTTCGGACACACTTGAAACATCTGAATGAACAAACTCTG	148355
OY		934	GACACACCATCTTTTCAGAGCTGTAACTCACCGCAAAGGCTCTGTGGCTTCATTCTTGAA	993
Db		148356	GACAGGCCATCTTTTAAAGACTGTAACTCACTCATGTAGAAGGCTTATGGGCTTCATTCTTGA	148415
OY		994	GTGACCAAGACAAGAACCCACCGGAAGAAACAATTCGACACAG	1040
Db		148416	GTGACCAAGACAAGAACCCACCGGAAGAAACAATTCGACACAG	148462
RESULT 26				
ACOL5806/C				
LOCUS	ACOL5806	190592 bp	DNA	linear HTG 26-MAY-2000
DEFINITION	Homo sapiens chromosome 11 clone RP11-420H22 map 11, WORKING DRAFT SEQUENCE, 10 unordered pieces.			
ACCESSION	ACOL5806			
VERSION	ACOL5806.3	GI:8096897		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 190592) Birren,B., Linton,L., Nusbaum,C. and Lande,E. Homo sapiens chromosome 11, clone RP11-420H22 Unpublished			
JOURNAL	2 (bases 1 to 190592)			
REFERENCE	Birren,B., Linton,L., Nusbaum,C., Lande,E., Allen,N., Anderson,M., Balwin,J., Batta,N., Beckerly,R., Boguslavsky,L., Bouckgalter,B., Brown,A., Brown,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hages,B., Harford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karlas,A., Klein,J., Lehoczy,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,S., Stojanovic,N., Subramanian,A., Talamas,J., Tenstade,S., Tirrell,A., Vassiliou,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.			
TITLE	Direct Submissions			
JOURNAL	Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
COMMENT	On May 26, 2000 this sequence version replaced gi:6692339. All repeats were identified using RepeatMasker: Smit,A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a>			
	----- Genome Center ----- Center: Whitehead Institute/ MIT Center for Genome Research Center code: MIBR Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a> Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a> ----- Project Information ----- Center project name: L2227 Center clone name: 420 H 22 ----- Summary Statistics ----- Sequencing vector: M13; M77815; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; Version 0.960731 Consensus quality: 184180 bases at least Q40 Consensus quality: 187115 bases at least Q30 Consensus quality: 188337 bases at least Q20 Insert size: 184000; agarose-FP Insert size: 189692; sum-of-contigs			

Quality coverage: 5.8 in Q20 bases; agarose-ip	
Quality coverage: 5.6 in Q20 bases; sum-of-contigs	
-----	
* NOTE: This is a 'working draft' sequence. It currently	
* consists of 10 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
* 1 3025: contig of 3025 bp in length	
* 3026 3125: gap of 100 bp	
* 3126 8902: contig of 5777 bp in length	
* 8903 9002: gap of 100 bp	
* 9003 16758: contig of 7756 bp in length	
* 16759 16858: gap of 100 bp	
* 16859 33120: contig of 16262 bp in length	
* 33121 33320: gap of 100 bp	
* 33321 49787: contig of 16567 bp in length	
* 49788 49887: gap of 100 bp	
* 49888 68295: contig of 18408 bp in length	
* 68296 68395: gap of 100 bp	
* 68396 93348: contig of 24953 bp in length	
* 93349 93448: gap of 100 bp	
* 93449 127332: contig of 33864 bp in length	
* 127333 127432: gap of 100 bp	
* 127433 156745: contig of 29213 bp in length	
* 156746 156745: gap of 100 bp	
* 156746 190592: contig of 33847 bp in length.	
* Location/Qualifiers	
FEATURES	1..190592
source	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="11"
	/map="11"
	/clone="RP11-420H22"
	/clone_lib="RPC1-11 Human Male BAC"
misc_feature	1..3025
	/note="assembly_fragment"
gap	3026..3125
	/estimated_length=100
misc_feature	3126..8902
	/note="assembly_fragment"
gap	8903..9002
	/estimated_length=100
misc_feature	9003..16758
	/note="assembly_fragment"
gap	16759..16858
	/estimated_length=100
misc_feature	16859..33120
	/note="assembly_fragment"
gap	33121..33220
	/estimated_length=100
misc_feature	33221..49787
	/note="assembly_fragment"
gap	49788..49887
	/estimated_length=100
misc_feature	49888..68295
	/note="assembly_fragment"
gap	68296..68395
	/estimated_length=100
misc_feature	68396..93348
	/note="assembly_fragment"
gap	clone_end:177
	vector_side:left"
gap	93349..93448
	/estimated_length=100
misc_feature	93449..127332
	/note="assembly_fragment"
gap	127333..127432
	/estimated_length=100

```
misc_feature 127433..156645
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:right"
gap           156646..156745
              /estimated_length=100
              156746..190592
              /note="assembly_fragment"
ORIGIN
Query Match      38.5%; Score 771; DB 14; Length 190592;
Best Local Similarity 86.3%; Pred. No. 3.5e-194;
Matches 921; Conservative 0; Mismatches 115; Indels 31; Gaps 5;

QY      1 AAAGGCTTGAATAGAGACATGCTTCAAACTCTTAATACCAACTCTGGAGTGGG 60
DB      99230 AAAGGCTTCCAAATAGAGACATCTTTCAAGCTCTTAATACCAAGCTCTGGAGTT-GGC 99172
QY      61 GACATGGCTTCTCCCTTTCTAGGTCCCTGAGAGAGCCATCTTCTAATAGTCGATTTGG 120
DB      99171 AACATGGCTTCTCCCTTTCTAGGTCCCAATGAGAGCCATCTTCTAATAGTCGATTTGG 99112
QY      121 GCCCTGATTTTAACTCTTGGTCAAAATTTGTTCTCTAGAGATGAGCCATCAAGCT 180
DB      99111 GCCCGATTTTAACTCTTGGTCAAAATTTGTTCTCTAGAGATGAGCCATCAAGCT 99052
QY      181 ACAGATGATCTTCAATATGTAACCCCAATGAGCTCACTAACAATCTTCTGTAGAGACC 240
DB      99051 ACAGATGATCTTCAATATGTAACCCCAATGAGCTCACTAACAATCTTCTGTAGAGACC 98995
QY      241 CCTGACCAACCCGCTGAGCCCTTTCAATGAGCTTAAAGAGCTCCCTCTGGAGAGACTAC 300
DB      98994 CCTGACCAACCCGCTGAGCCCTTTCAATGAGCTTAAAGAGCTCCCTCTGGAGAGACTAC 98935
QY      301 CACTGAGAGGCCCTTTCTTACCCCTATCCAGAGAGAGTACTAGAGCGGTATC-GCC 359
DB      98934 AACCTGAGAGGCTCTTTCTTTGCCCCCTATCCAGAGAGAGTACTAGAGCGGTATCACC 98875
QY      360 AATATCCCAACAGAGCTGGGGGTCTCTGTTTGAAGGGGGATTGAAGGTGAAGCCAGCT 419
DB      98874 AATATCCCAACAGAGCTGGGGGTCTCTGTTTGAAGGGGGATTGAAGGTGAAGCCAGCT 98815
QY      420 GGGCTTCTGGGTCAAGGTGGGGAACTTGGAGAACTTTTGTGCTAGCTAAAGGATTGTAAT 479
DB      98814 GGGCTTCTGGGTCAAGGTGGGGAACTTGGAGAACTTTTGTGCTAGCTAAAGGATTGTAAT 98755
QY      480 GACCAATGAGACTCTGTGTCTAGCTAAAGATTGTAATGC-----522
DB      98754 ACACCAATGAGCTATGTGTCTAGCTAAAGATTGTAATGACCAATGAGACTCTGT 98695
QY      523 -----ACCAATGAGACTCTGTGTCTAGCTAAAGATTGTAATGC-----574
DB      98694 AAAAAAGCAATGAGACTCTGTGTGTAAATGACCAATGAGAGTGGGGGAGGAGCA 98635
QY      575 AATAAGGGAGTAAATACTGGCCACCCGAGCCAGCACTGGCAACCCCTGGGGCCCCCTTC 634
DB      98634 AATAAGGGAGTAAATACTGGCCACCCGAGCTGAGCACTGGCAACCCCTGGGGCCCCCTTC 98575
QY      635 CACACTGTGAAGCTTTGTTGTTGCTCTTCAATTAATCTTGTGCTGCTCATTTT 694
DB      98574 CACTGTGTGAAGCTTTGTTGTTGCTCTTCAATTAATCTTGTGCTGCTCATTTT 98515
QY      695 TGTGTTCACACTTCTTTTATGAGCTTAAACTCACTGAGAGGGTCTTGGCTTCAATTC 754
DB      98514 TGTGTTCACACTTCTTTTATGAGCTTAAACTCACTGAGAGGGTCTTGGCTTCAATTC 98455
QY      755 TGAAGTCAAC-AGACACGAGAACCCACTGGAAGGAACAAGAACTCCGAGTGTGGCTT 813
DB      98454 TGAAGTCAACGAGAACCCACTGGAAGGAACAAGAACTTGAAGTGTGGCTT 98395
QY      814 TAAGAGCTTAAACTCACTGAGAGCTTGAAGCTTCACTCTGTAAGTCAAGTGAAGCA 873
DB      98394 TAAGAGCTTAAACTCACTGAGAGCTTGAAGCTTCACTCTGTAAGTCAAGTGAAGCA 98335
```

```
QY      874 CAAACCCACGAAAGGAAGAACTCTGACACACTGAAATATCTGAAGAAACAACCTCCA 933
DB      98334 CAAACCCACTGGAAGGAAGAAATTTTGGACACACTTGAACATCTGAATGAACAACCTCG 98275
QY      934 GACACACACTTTTCTAGAGCTGTAAACTCACTGACCGCAAGGGTCTGTGGCTTCAATTTGA 993
DB      98274 GACACACCACTTTTAAAGACTGTAACTCACTGAGAGGGTCTATGGGTTCAATTTGA 98215
QY      994 GTGACGAAGACCAAGAACCCACCGAAGGAACAATTCAGACACAG 1040
DB      98214 GTGACGAAGACCAAGAACCCACCGAAGGAACAATTCGAGACAG 98168

RESULT 27
AC024401 168298 bp DNA linear HTG 04-MAY-2001
LOCUS     Homo sapiens chromosome 2 clone RP11-488M11 map 2, WORKING DRAFT
DEFINITION
ACCESSION AC024401.2 GI:9994150
VERSION    HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE 1 (bases 1 to 168298)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beza,F., Boguslavsky,L.,
            Bourkhalter,B., Brown,A., Burkett,G., Campobiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
            Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fensholt,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hages,B., Heatford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Laroque,K., Lechoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGuirk,A., McKernan,K., McPheters,R., Meldrum,J.,
            Menus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tittell,A.,
            Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
            Direct Submission
            Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Sep 8, 2000 this sequence version replaced gi:7108197.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            -----
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: W1BR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence submissions@genome.wi.mit.edu
            ----- Project Information
            -----
            Center project name: 16777
            -----
            Center clone name: 488.M.11
            ----- Summary Statistics
            Sequencing vector: M13, M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.96071
            Consensus quality: 163206 bases at least Q40
```

Consensus quality: 166077 bases at least Q30  
Consensus quality: 166822 bases at least Q20  
Insert size: 168000; agarose-fp  
Insert size: 167098; sum-of-contigs  
Quality coverage: 5.3 in Q20 bases; agarose-fp  
Quality coverage: 5.3 in Q20 bases; sum-of-contigs

\*\*\*\*\*  
NOTE: This is a 'working draft' sequence. It currently  
consists of 13 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 19212: contig of 19212 bp in length  
\* 19213 19312: gap of 100 bp  
\* 19313 20607: contig of 1295 bp in length  
\* 20608 20707: gap of 100 bp  
\* 20708 24577: contig of 3870 bp in length  
\* 24578 24677: gap of 100 bp  
\* 24678 29497: contig of 4820 bp in length  
\* 29498 29597: gap of 100 bp  
\* 29598 33210: contig of 3613 bp in length  
\* 33211 33310: gap of 100 bp  
\* 33311 38616: contig of 5306 bp in length  
\* 38617 38716: gap of 100 bp  
\* 38717 47747: contig of 9031 bp in length  
\* 47748 47847: gap of 100 bp  
\* 47848 59600: contig of 11753 bp in length  
\* 59601 59700: gap of 100 bp  
\* 59701 72620: contig of 12920 bp in length  
\* 72621 72720: gap of 100 bp  
\* 72721 99437: contig of 26717 bp in length  
\* 99438 99537: gap of 100 bp  
\* 99538 126237: contig of 26700 bp in length  
\* 126238 149961: contig of 23624 bp in length  
\* 149962 150062: gap of 100 bp  
\* 150062 168298: contig of 18237 bp in length.

## FEATURES

source

1. 168298  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-488M11"  
/clone\_1fb="RPC1-11 Human Male BAC"  
1. 19212  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
19213..19312  
/estimated\_length=100  
19313..20607  
/note="assembly\_fragment"  
20608..20707  
/estimated\_length=100  
20708..24577  
/note="assembly\_fragment"  
24578..24677  
/estimated\_length=100  
24678..29497  
/note="assembly\_fragment"  
29498..29597  
/estimated\_length=100  
29598..33210  
/note="assembly\_fragment"  
33211..33310  
/estimated\_length=100  
33311..38616  
/note="assembly\_fragment"

## ORIGIN

Query Match 38.0%; Score 761.2; DB 14; Length 168298;  
Best Local Similarity 88.0%; Pred. No. 1.4e-191;

Matches 928; Conservative 0; Mismatches 98; Indels 26; Gaps 8;

QY	1	AAAGGCTTGTGAATATGACAAATGCTTTCAACTTTATACCACTTGTGAGTGGG	60
DB	91219	AAAGGCTTGTGAATATGACAAATGCTTTCAACTTTATACCACTTGTGAGTGGG	91278
QY	61	GACATGGCTTCCCTTTCTAGTCTGAGCAGCATCTGCTAATGTGCACTTTGG	120
DB	91279	AACATGGCTTCCCTTTCTAGTCTGAGCAGCATCTGCTAATGTGCACTTTGG	91338
QY	121	GGCCTGATTTTAACTCTTGCTCAAAATTTGTTTCTTAGAGTGAAGCCATCAAGCT	180
DB	91339	GGCCTGATTTTAACTCTTGCTCAAAATTTGTTTCTTAGAGTGAAGCCATCAAGCT	91398
QY	181	ACAGATGATCTTAAATGATTAACCCCAATGAGCTAATCACTTGTGAGAGCC	240
DB	91399	ACAGATGATCTTAAATGATTAACCCCAATGAGCTAATCACTTGTGAGAGCC	91458
QY	241	CTTGACCGACCGCGCTGCGCTTTCAATGAGCCCTTAAGAGTCCCTTGAGGACACTAC	300
DB	91459	CTTGAGGTGACCGCGCTGCGCTTTCAATGAGCCCTTAAGAGTCCCTTGAGGACACTAC	91518
QY	301	CACGTGAGGCGCCCTTCTTCAACCCCTATCCAGCAGAGTAGCTACAGCGCATCG-CC	359
DB	91519	AACGTGAGGCGCCCTTCTTCAACCCCTATCCAGCAGAGTAGCTACAGCGCATCG-CC	91578
QY	360	AAATCCCAACAGCAGTGGGGTGTCTGTTGAGAGGGGGATTGAGAGTGAAGCCAGCT	419
DB	91579	AAATCCCAACAGCAGTGGGGTGTCTGTTGAGAGGGGGATTGAGAGTGAAGCCAGCT	91638
QY	420	GGGCTT-CTGGGTACAGTGGGACCTTGAGAACTTTTGTCTAGCTTAAGGATGTGAAA	478
DB	91639	GGAATTCCTGGGTACAGTGGGACCTTGAGAACTTTTGTCTAGCTTAAGGATGTGAAA	91697
QY	479	TGACCAATGAGCAGTGTGTCTAGCTAAGATTTGTAATGACCAATCAGCAGCTCG	538
DB	91698	CGACCAATGAGCAGCTGTG-----TAAATGACCAATCAGTGTCTCG 91741	
QY	539	TAAATGACCAATCAGCAGTGTGTGGCGGGTCAAAATGAGGATTAATAAATCTGGCCAC	598

```

Db      91742 TAAAAAGACTAATGACAGAAAGT-GGCGGGCCAAATTAAGGAAATAAAGCTGGCGCA- 91799
QY      599 CCGAGCGAGGAGGGAACCCACTCGG-----GTCCCTTCCACACTGGGAAGCTTTG 652
Db      91800 CCGAGCGAGGAGGGAACCCACTCGGCTGGTCCAGTCCCTTCCACACTGGGAAGCTTTG 91859
QY      653 TTCTTTGGCTCTTCACAATTAATCTTGTCTGCTCTCAATTTTGTGTCCACTACCTTT 712
Db      91860 TTCTTTGGCTCTTCACAATTAATCTTGTCTGCTCTCAATTTTGTGTCCACTACCTTT 91919
QY      713 ATGAGCTGTAACTCACTGCGAGGGTCTGTGGCTTCATTCTGAAGTCA-ACAGACCAC 771
Db      91920 ATGAGCTGTAACTCACTGCTGAGGGTCTGGCTTCACTCAAGTCAAGTGAAGACAC 91979
QY      772 GAACCCACTGGAAGAAACAAAGAACTCCCATATGCTGCTTTAAAGCTGTAACTCA 831
Db      91980 AAACCCACCAAGGGAACAAACAACTCTGTATGACACCACTTTAAGAGCTGTAACTCA 92039
QY      832 CTGCGAAGCTCTGACCTTCACTCTGTAAGTCAAGTGAAGACCAACCCACGAAGGAA 891
Db      92040 CTGCGAAGCTCTGCTGCTCTCACTCTGTAAGTCAAGTGAAGACCAACCCACGAAGGAA 92099
QY      892 GAAACTCTGGAACACACCTGTAATATCTGAAGAAACAACTCCAGACACACCATCTTTTCA 951
Db      92100 GAAACTCTGGAACACATCTGAACATCTGAAGAAACAACTCCAGACACATCTTTTAA 92159
QY      952 GCTGTAACTCACTGCGGAGGGTCTGTGGCTTCATTCTTGAAGTCAAGACCAAGAAC 1011
Db      92160 ACTGTAACTCACTGCGGAGGGTCTGTGGCTTCATTCTTGAAGTCAAGACCAAGAAC 92219
QY      1012 CCACCGGAAGGAACAAATTCAGACACAGTAA 1045
Db      92220 CCACCGGAAGGAACAAATTCAGACACAGTAA 92253

RESULT 28
AC013406 190018 bp DNA linear PRI 21-Apr-2005
LOCUS Homo sapiens BAC clone RP11-341F20 from 2, complete sequence.
AC013406
VERSION AC013406.7 GI:10312292
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 190018)
Kalicki, J., Ureta, M. and Johnson, D.
Unpublished (2001)
2 (bases 1 to 190018)
Waterston, R.H.
Direct Submission
Submitted (09-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 190018)
Waterston, R.H.
Direct Submission
Submitted (27-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 190018)
Waterston, R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 190018)
Wilson, R.K.
Direct Submission
Submitted (21-APR-2005) Genome Sequencing Center, Washington

```

University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 On Sep 27, 2000 this sequence version replaced gi:9945536.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@wustl.wustl.edu](mailto:submissions@wustl.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0341F20  
 -----

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Caranese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBAC3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-634D21; the clone sequenced to the right is RP11-219B1. Actual start of this clone is at base position 1 of RP11-341F20; actual end is at base position 190018 of RP11-341F20.

## FEATURES

source location/Qualifiers  
 1..190018  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /clone\_1fb="RP11-341F20"  
 /clone\_1fb="RP11-11"  
 72290..72518  
 /note="CpG\_island (%GC=62.0, o/e=0.86, #CpGs=19)"

## misc\_feature

## ORIGIN

Query Match 38.0%; Score 761.2; DB 8; Length 190018;  
 Best Local Similarity 88.0%; Pred. No. 1,4e-191;  
 Matches 928; Conservative 0; Mismatches 98; Indels 28; Gaps 8;

```

QY      1 AAAGGCTTGAATGACAGCAATGCTTTCAACTGTATACCAACTGTGAGTTGGG 60
Db      20888 AAAGGCTTGTGAATGACAGCAATGCTTTCAAACTTTATACCAACTGTGAGTTGGG 20947
QY      61 GACATGCTTCTCCCTTTCTAGGCTCTGTGAACGCCATCTTGCTAATAGTGCATTTGG 120
Db      20948 AACATGCTTCTCCCTTTCTAGGCTCTGTGAACGCCATCTTGCTAATAGTGCATTTGG 21007
QY      121 GCCGTATATTTTAACCTTGTGTAATTTGTTTCTGTAGAGTGAAGCCATGAAGT 180
Db      21008 GCCGTATATTTTAACCTTGTGTAATTTGTTTCTGTAGAGTGAAGCCATGAAGT 21067
QY      181 ACAGATGATCTTAAATGAATGAACCCAAATGAAGTCACTAACTTGTGTAAGACC 240

```

```

Db      21068 ACAGATGATCTTAAACAATGAAACCCCAATTAAGCTCAGTAACTTCTACCGAGAC 21127
Qy      241 CCTGACCGACCGCGTGGCCCTTTCAATGAGCCTTAAGAGCTCCCTCTGAGAGCACTAC 300
Db      21128 CTTGAGATGACCGCGTGGCCCTTTCACTGAGCTTAAGAGTTCCCTCTGAGAGCACTAC 21187
Qy      301 CACTGAGGAGCCCTTTCTTCAACCCATCTCAGACGAGAGTACTACAGCGGTATCG-CC 359
Db      21188 AACTGAGGAGCCCTTTCTTCAACCCATCTCAGACGAGAGTACTAGTGTATCGTCC 21247
Qy      360 AATCCCAACAGACGCTGGGTCTCTGTTTGAAGGGGAGTTGAGAGTGAAGCCAGCT 419
Db      21248 AATTCCTCAACAGAGTGGGTCTCTGTTTGAAGGGGAGTTGAGAGTGAAGCCAGCT 21307
Qy      420 GAGGCTT-CTGGGTCAAGTGGGAGCTTGAAGACTTTTGTCTAGCTTAAAGATTGTA 478
Db      21308 GAGCTTCTGGGTCAAGTGGGAGCTTGAAGACTTTTGTCTAGCTTAA-CTTTGTAA 21366
Qy      479 TGCACCAATCAGACCTCTGTGTCTAGCTTAAAGATTGTAAGTGAAGCCAGCTCTG 538
Db      21367 CGCACCATCAGACCTCTG-----TAAATGCAACCAATCAGTGTCTG 21410
Qy      539 TAAATGACCAATCAGACGAGATGTGGCGGGCTCAATTAAGGAGTAAATAATGGCCAC 598
Db      21411 TAAATGACCAATCAGACCAAGT-GGCGGGCAATTAAGGAGTAAATAATGGCCAC 21468
Qy      599 CCGAGCAGACGAGTGGCAACCACTCG3-----GTCCCTTCCACTGTGAAGCTTGG 652
Db      21469 CCGAGCAGACGAGTGGCAACCTCGTGGTCCAGTCCCTTCCACTGTGAAGCTTGG 21528
Qy      653 TTCTTTGCTCTTCAACATTAATCTGTGCTGCTCATCTTTGTGTCCACTTACCTTT 712
Db      21529 TTCTTTGCTTTTCAACATTAATCTGTGCTGCTCATCTTTGTGTGCTGCTGCTT 21588
Qy      713 ATAGCTGTAACTCACTAGCGAGGCTGTGCTTATCTCGAAGTCA-ACAGACCAC 771
Db      21589 ATAGCTGTAACTCACTAGCGAGGCTGTGCTTATCTCGAAGTCAAGTGAAGCAC 21648
Qy      772 GAACCACTGGAAGGAACAAGAACTCCGATGCTGCCTTTAAGAGCTGTAACTCA 831
Db      21649 AAACCCACGAGAGGAACAACACTTGATGACCACTTTAAGAGCTGTAACTCA 21708
Qy      832 CTGCGAAGCTCTGCGCTTCACTCTGTAAGTCAAGTGAAGCAACCAACCAAGAA 891
Db      21709 CTGCGAAGCTCTGCGCTTCACTCTGTAAGTCAAGTGAAGCAACCAACCAAGAA 21768
Qy      892 GAAACTCTGACACACTGTAATATCTGAAGGAACAACCTCAGACACCACTTTTGA 951
Db      21769 GAAACTCTGACACACTGTAATATCTGAAGGAACAACCTCAGACACCACTTTTGA 21828
Qy      952 GCTGTACACTCAACCGAAGGCTGTGCTTCACTTGAAGTCAAGACCAAGAC 1011
Db      21829 ACTGTAACTCAACCGAAGGCTGTGCTTCACTTGAAGTCAAGACCAAGAC 21888
Qy      1012 CCACCGAAGGAACAATTCAGACACAGTAGGA 1045
Db      21889 CCACCGAAGGAACAATTCAGACACAGTAGGA 21922

```

```

RESULT 29
AC098972      145993 bp      DNA      linear      PRI 22-AUG-2002
LOCUS      Homo sapiens chromosome 3 clone RP11-197J13, complete sequence.
DEFINITION      AC098972 AC031986
ACCESSION      AC098972.2 GI:22417369
VERSION      HTG.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE      1 (bases 1 to 145993)
AUTHORS      Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,

```

```

TITLE      Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
JOURNAL      Haugen,E.D.
REFERENCE      Direct Submission
AUTHORS      Unpublished
TITLE      2 (bases 1 to 145993)
JOURNAL      Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
REFERENCE      Direct Submission
AUTHORS      Submitted (07-NOV-2001) Genome Center, University of Washington,
TITLE      Box 352145, Seattle, WA 98195, USA
JOURNAL      3 (bases 1 to 145993)
REFERENCE      Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
AUTHORS      Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
TITLE      Haugen,E.D.
JOURNAL      Direct Submission
REFERENCE      Submitted (22-AUG-2002) Genome Center, University of Washington,
AUTHORS      Box 352145, Seattle, WA 98195, USA
COMMENT      On Aug 22, 2002 this sequence version replaced gi:16756284.

```

```

Center: University of Washington Genome Center
Center Code: UMG
Web site: http://www.genome.washington.edu
Contact: umgchgs@u.washington.edu
Drafting Center: BCM
Project Information
Center project name: chr-3
Center clone name: RP11-197J13 (bc0295)

```

```

Summary Statistics
Sequencing vector: unknown; 45% of reads
Sequencing vector: plasmid; 55% of reads
Sequencing vector: plasmid; 108752; 0% of reads
Chemistry: Dye-terminator ET; 80% of reads
Chemistry: Dye-terminator Big Dye; 20% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 145920 bases at least Q40
Consensus quality: 145993 bases at least Q30
Insert size: 145993; sum-of-contrigs
Quality coverage: 11.4x in Q20 bases; sum-of-contrigs

```

```

Overlapping Sequences:
5': RP11-655N11 (UMGC:bc0519) AC099057, 55267-bp overlap
3': RP11-638F19 (UMGC:bc0505) AC107622, 53874-bp overlap

```

```

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

```

```

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

```

#### Sequence Validation:

```

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

```

Bg111

Nst1

HindIII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
5583	5542	16518	16670	2251	2300
2067	2054	6048	6057	6382	6424
6916	7112	1968	1944	512	<800
1415	1397	8820	8839	449	<800
3539	3505	10950	10760	5313	5239
2898	2922	1125	1128	934	878
1716	1701	18381	18445	2061	2030
2858	2922	2937	3016	865	878
343	<800	150	<800	1557	1601
263	<800	462	<800	4158	4060
2502	2527	259	<800	683	<800
2325	2304	4527	4496	623	<800
2166	2161	380	<800	3684	3797
2301	2304	372	<800	3569	3610
3701	3698	545	<800	712	<800
1139	1148	81	<800	1265	1264
776	766	4438	4496	1030	994
1761	1701	5428	5485	2998	3112
5888	5931	160	<800	6492	6424
768	766	4957	4932	3155	3112
6502	6578	6821	6819	5926	5890
3345	3369	1025	1008	2538	2505
4223	4226	1307	1298	2324	2200
12029	11859	962	1008	562	<800
3455	3505	245	<800	2035	2030
603	<800	12092	11821	992	994
5223	5224	696	<800	891	878
2691	2683	769	757	412	<800
4386	4392	1278	1298	754	<800
3447	3505	514	<800	2731	2774
13647	13699	1636	1619	1281	1264
291	<800	9574	9578	77	<800
747	766	454	<800	1792	1868
9341	9385	1562	1619	3553	3478

QY	1	AAAGCTTCGAAATGAGCAATGCTTTCAAACCTTATACCACTCTGAGTTGGC	60
DB	14358	ACAGGCTTCGAAATGAGCAATGCTTTCAAACCTTATACCAATGCTGAGTTGGC	143417
QY	61	GACATGCTTCCTCCCTTTCAGTCTGTGACAGCCATCTTGTAATGTCATTTGG	120
DB	143418	AACATGCTTCCTCCCTTTCAGTCTGTGACAGCCATCTTGTAATGTCATTTGG	143477
QY	121	GCCCTGATTTTAACTCTTGTCAAATTTGTTTCTTAGGATGAGGCCATCAACT	180
DB	143478	GACCTGATTTTAACTCTT-TTCAAGTTTGTTCCTAGATGAGGCCATCAAGCT	143536
QY	181	ACAGATGCTTACAAATGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG	240
DB	143537	ACAGATGCTTACAAATGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG	143596
QY	241	CCTGACCAACCGCTGCTTTCATGAGCTTAAGAGCTCCCTCTGAGACACTAC	300
DB	143597	CTTGACCAACCGCTGCTTTCATGAGCTTAAGAGCTCCCTCTGAGACACTAC	143656
QY	301	CACCTGAGGCGCTTCTTCAACCTTATCCAGAGAACTAGCTACAGGCTATC-GCC	359
DB	143657	AACCTGAGGCGCTTCTTCAACCTTATCCAGAGAACTAGCTATCAGGCTATCACC	143716
QY	360	AAATCCCAACAGAGCTGAGGCTGCTGTTTGGAGGAGGATGAGGCTGAAGCCAGCT	419



Db 143717 GATTCCTCAACAGCAGTTAGGCTGCTCTTTAGAGGAGAGACTGAGAGGTGAACCCGGCT 143776  
QY 420 GGGCTTCTGGGTGAGGTGGGAGCTTGAGAACTTTGTGTCTAGTAAAGATTGTAAAT 479  
Db 143777 GGGCTTCTGGGTGAGGTGGGAGCTTGAGAACTTTGTGTCTAGTAAAGATTGTAAAT 143836  
QY 480 GCACCAATGAGACTCTGTGTCTAGCTAAAGATTGTAAATGAGCAATCAGACT- 535  
Db 143837 GCACCAATGAGACTCTGTGTCTAGCTAAAGATTGTAAATGAGCAATCAGACTCTGT 143896  
QY 536 ----- 535  
Db 143897 AAAAATGACCAATCAGACTCTGTAAATGGGTCAATGACACTCTGTAAATGAGCA 143956  
QY 536 -----CTGTAAATGAGCAATCAGAGAGATGGGCGGGGTCAATTAAGGAGTA 586  
Db 143957 ATCAGCCCCCTGTAAATGAGCAATCAGAGATGGGCGGGGTCAATTAAGGAGTA 144016  
QY 587 AAAAATGAGCAATCAGAGAGATGGGCGGGGTCAATTAAGGAGTA 646  
Db 144017 AAAAGTGGGCACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144076  
QY 647 GCTTTGTTCTTTTGTCTTCTTCAATTAATTTGTCTGTCTGCTATCTTTGTCTTCAACT 706  
Db 144077 GCTTTATCTCTTGTCTTCTTCAATTAATTTGTCTGTCTGCTATCTTTGTCTTCAACT 144136  
QY 707 ACCTTATGAGTGTAACTCACTGAGAGAGTCTGTGCTTCAATTTCTTGAAGTCA-ACA 765  
Db 144137 ACCTTATGAGTGTAACTCACTGAGAGAGTCTGTGCTTCAATTTCTTGAAGTCA 144196  
QY 766 GACCAAGAACCCAGTGGAG 825  
Db 144197 GACCAAGAACCCAGTGGAG 144256  
QY 826 CACTCACTGGAGAGTCTGAGCTTCACTCTTGAAGTCAAGTCAAGAGAGAGAGAGAGAGAG 885  
Db 144257 CACTCACTGGAGAGTCTGAGCTTCACTCTTGAAGTCAAGTCAAGAGAGAGAGAGAGAG 144316  
QY 886 AAGGAAGAACTCTGAG 945  
Db 144317 AAGGAAGAACTCTGAG 144376  
QY 946 TTGAGAGCTTAACTCACTCAGCAGAGAGTCTGTGCTTCAATTTCTTGAAGTCAAGAGAGAG 1005  
Db 144377 TTGAGAGCTTAACTCACTCAGCAGAGAGTCTGTGCTTCAATTTCTTGAAGTCAAGAGAGATC 144436  
QY 1006 AAGAACCCAGCAG 1041  
Db 144437 AAGAACCCAGCAG 144472

RESULT 30  
AC107622 158508 bp DNA linear PRI 29-MAR-2002  
LOCUS Homo sapiens chromosome 3 clone RP11-638F19, complete sequence.  
DEFINITION AC107622  
ACCESSION AC107622 GI:19807856  
VERSION AC107622.2 GI:19807856  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE 1 (bases 1 to 158508)  
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 158508)  
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (24-JAN-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

REFERENCE 3 (bases 1 to 158508)  
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAR-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
COMMENT On Mar 29, 2002 this sequence version replaced gi:18308300.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: <http://www.genome.washington.edu>  
Contact: [uwgchgs@u.washington.edu](mailto:uwgchgs@u.washington.edu)  
----- Project Information  
Center project name: chr-3  
Center clone name: RP11-638F19 (bc0505)  
----- Summary Statistics  
Sequencing vector: plasmid; 94% of reads  
Sequencing method: Dye-terminator; 47% of reads  
Chemistry: Dye-terminator; 47% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 158394 bases at least Q40  
Consensus quality: 158506 bases at least Q30  
Insert size: 158508; sum-of-contigs  
Quality coverage: 7.8x in Q20 bases; sum-of-contigs

-----  
Overlapping Sequences:  
5': RP11-197013 (UWGC:bc0295) AC098972  
3': RP11-669C19 (UWGC:bc0523) AC116096  
-----

Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

#### Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HandII	BglII	EcoRI
SedDerMap FngRPrint	SedDerMap FngRPrint	SedDerMap FngRPrint
-----	-----	-----
3990	3966	10712
-----	-----	-----
6382	6425	2067
-----	-----	-----
512	<800	10737
-----	-----	-----
449	<800	291
-----	-----	-----
		705
-----	-----	-----
		<800
-----	-----	-----

8252	8252	747	751	688	<800	301	<800	5242	5139	2171	2340
779	<800	9340	9505	7499	7596	1241	1230	2733	2899	10020	9872
1616	1638	333	<800	7576	7596	1063	1060	769	<800	110	<800
1757	1765	1437	1405	589	<800	1648	1638	3234	3217	6884	6817
3387	3427	863	929	696	<800	1986	1886	3555	3472	3504	3531
385	<800	406	<800	205	<800	131	<800	13420	12963	1242	1192
2786	2900	163	<800	12176	11807	2540	2581	3394	3472	1371	1347
4094	3966	214	<800	717	<800	8828	8870	3073	3217	210	<800
3291	3244	3057	3217	645	<800	941	934	6117	6509	291	<800
2307	2306	1347	1303	2251	2340	81	<800	1630	1579	5537	5595
1214	1230	2330	2458	1982	1998	5708	5610	14871	15751	675	<800
1115	1133	540	<800	3007	3036	4925	4826	542	5595	5542	5595
2839	2900	347	<800	199	<800	1890	1886	17062	17584	17062	17584
2115	2051	282	<800	276	<800	6562	6425	90	<800	90	<800
321	<800	424	<800	1053	1036	2236	2306	6028	6011	6028	6011
3973	3966	897	929	486	<800	463	<800	1561	1494	1561	1494
6441	6425	1792	1832	12153	11807	11675	11738	1164	1192	1164	1192
1899	1886	185	<800	258	<800						
39	<800	546	<800	687	<800						
4367	4285	2383	2562	3290	3295						
2813	2900	1580	1579	442	<800						
1063	1060	4330	4631	2289	2340						
911	838	4658	4882	3697	3745						
557	<800	1189	1175	2385	2519						
5965	5999	4503	4702	2750	2885						
2195	2306	3658	3798	241	<800						
10103	10081	942	974	2017	2154						
687	<800	10460	10092	5854	5595						
1360	1371	6074	5775	1432	1494						
791	808	2921	2899	860	859						
3439	3427	1960	2113	1245	1192						
1568	1550	2684	2618	121	<800						
10038	10081	1647	1579	474	<800						
2481	2581	228	<800	1552	1494						
2858	2900	4411	4702	2761	2885						
3849	3808	6484	6845	769	753						

Query Match  
Best Local Similarity 83.7%; Score 754.4; DB 8; Length 158508;  
Matches 934; Conservative 0; Mismatches 106; Indels 76; Gaps 4;

QY 1 AAGGCTTGAATGACATGCTTTCAAACTTTATACCACTGTGAGTTGGC 60  
DB 51237 ACAGGCTTGAATGACATGCTTTCAAACTTTATACCACTGTGAGTTGGC 51296  
QY 61 GACATGCTTCCCTTTAGTCTGAGAGCACTTTCTAATATATGCAATTGG 120  
DB 51297 AACATGCTTCCCTTTAGTCTGAGAGCACTTTCTAATATATGCAATTGG 51356  
QY 121 GCCCTGATTTTAACTCTTGTCAAAATTTGTTCTAGAGTGAAGGCACTCAAGCT 180  
DB 51357 GACCTGATTTTAACTCTCT-TTCAAGTTGTTCTCTAGAGTGAAGGCACTCAAGCT 51415  
QY 181 ACAGATGATCTTAAATGTAATGACCCCAATGAGCTCACTAACAATCTTGTGAGAGCC 240  
DB 51416 ACAGATGATCTTAAATGTAATGACCCCAATGAGCTCACTAACAATCTTGTGAGAGCC 51475  
QY 241 CTTGAGACCAACCCGCTGCTTTCAATGAGGCTTCAAGGCTTCCCTCTGAGAGCACTAC 300  
DB 51476 CTTGAGACCAACCCGCTGCTTTCAATGAGGCTTCAAGGCTTCCCTCTGAGAGCACTAC 51535  
QY 301 CACTGAGAGGCGCTTCTTCAACCCCTATCCAGAGAGTAAGAGGCTATC-GGC 359  
DB 51536 AACTGAGAGGCGCTTCTTCAACCCCTATCCAGAGAGTAAGAGGCTATC-GGC 51595  
QY 360 AATCCCAACAGCAGTGGGCTGCTGTTTGGAGGGGAGATTGAGAGTGAAGCAAGCT 419  
DB 51596 GATTCCCAACAGCAGTGGGCTGCTGTTTGGAGGGGAGATTGAGAGTGAAGCAAGCT 51655  
QY 420 GGGCTTCTGGGCTGAGTGGGAGCTTGGAGAACTTTGTGTCTGCTAAAGATTTGTAAT 479  
DB 51656 GGGCTTCTGGGCTGAGTGGGAGCTTGGAGAACTTTGTGTCTGCTAAAGATTTGTAAT 51715  
QY 480 GCACCATGACACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGCACT---- 535

Db	51216	GCACCAATCAAGCGCTGTGTCTTAGCTAAAGATTGCAAAAGCAACCAATCACTCTGT	51775
QY	536	-----	535
Db	51776	AAAATGCACCAATCAAGCACTCTGTAAATGGGTCAATGACGACTCTGTAAATGGACCA	51835
QY	536	-----CTGTAAATGGACCAATAGAGAGATGTGGGCGGGGTCAAATTAAGGAGTA	586
Db	51836	ATCAGCCCCCTGTAAATGGAACAATAGAGATGTGGGCGGGGTCAAATTAAGGAGTA	51895
QY	587	AAAATGGGCAACCCGAGCCAGAGTGGCAACCCACTGGGTCCCTTCACACTGTGGAA	646
Db	51896	AAAGCTGGCACTGAGACCGAGAGTGGCAACTGTAGAGGTCCCAATTCACGCTGTGGAA	51955
QY	647	GCTTGTGTTCTTTGCTCTTCAATTAATTTTGCTGCTCACTTTCTTGTCTCACT	706
Db	51956	GCTTATTTCTTTGCTCTTCAATTAATTTTGCTGCTCACTTTTGTTGCTCACT	52015
QY	707	ACCTTATGAGCTGTAACTCTCACTGGAAGGTCTGGGTCTTATTCGGAAGTCA -ACA	765
Db	52016	ACCTTATGAGCTGTAACTCTCACTGGAAGGTCTGGGTCTTATTCGGAAGTCA	52075
QY	766	GACCAAGCAACCCACTGGAAGGAACAAGAATCCGATGTGCTCCTTTAAGAGCTGTAA	825
Db	52076	GACCAAGCAACCCACTGGAAGGAACAAGAATCCGATGTGCTCCTTTAAGAGCTGTAA	52135
QY	826	CACCTCACTGCGAAGCTCTGCGAGTTCACTCTCTGAAGTCAAGTGAAGCAACCAACCCAG	885
Db	52136	CACCTCACTGCGAAGGTCTGTGGCTTCACTCTCTGAAGTCAAGTGAAGCAACCAACCCAG	52195
QY	886	AAGGAAGAAATCTGTGAGACACCTGGAATATCTGAAGGAACAAACTCCAGACACACATCT	945
Db	52196	AAGGAAGAAATCTGTGAGACACCTGGAATATCTGAAGGAACAAACTCCAGATACACATCT	52255
QY	946	TTCAAGAGCTGTAACTCTCAACCGAAGGCTGTGGCTTCACTTTCTGGAAGTCAAGACAC	1005
Db	52256	TTAAGAACTGTAACTCTCAACCATAGAGGTCCGTGGCTTCACTTTCTGTAAGTCAAGACAGATC	52315
QY	1006	AAGAACCCACCGAAGGAACAAATTCGAGACACAGT	1041
Db	52316	AAGAACCCACCTGGAAGGAACCAATTCGAGACACACT	52351
RESULT 31			
LOCUS	BX571893/C	94768 bp DNA linear HTG 25-JUL-2003	
DEFINITION	Homo sapiens chromosome 6 clone DASS-75011, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.		
ACCESSION	BX571893		
VERSION	BX571893.1 GI:33284734		
KEYWORDS	HTG; HTGS; PHASE1.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 94768)		
REFERENCE	Sims,S.		
AUTHORS	Direct Submission		
TITLE	Submitted (24-JUL-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CH10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
JOURNAL	Genome Center		
COMMENT	Center: Wellcome Trust Sanger Institute		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: humquerry@sanger.ac.uk		
	Project Information		
	Center project name: bss75011		
	Summary Statistics		
	Assembly program: XGAP4; version 4.5		
	Chemistry: Dye-terminator; 100% of reads		
	Consensus quality: 92859 bases at least Q40		

		Consensus quality: 93559 bases at least Q30	
		Consensus quality: 94137 bases at least Q20	
		Insert size: 99204; 1.9% error; agarose-fp	
		Quality coverage: 4.88x in Q20 bases; sum-of-contigs	
		coverage: 4.81x in Q20 bases; agarose-fp	
		-----	
		* NOTE: This is a 'working draft' sequence. It currently	
		* consists of 4 contigs. The true order of the pieces	
		* is not known and their order in this sequence record is	
		* arbitrary. Gaps between the contigs are represented as	
		* runs of N, but the exact sizes of the gaps are unknown.	
		* This record will be updated with the finished sequence	
		* as soon as it is available and the accession number will	
		* be preserved.	
		1 31974: contig of 31974 bp in length	
		31975 32074: gap of 100 bp	
		32075 51357: contig of 19283 bp in length	
		51358 51457: gap of 100 bp	
		51458 83889: contig of 32432 bp in length	
		83880 83989: gap of 100 bp	
		83990 94768: contig of 10779 bp in length.	
		Location/Qualifiers	
		1..94768	
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/BD_xref="taxon:9606"	
		/Chromosome="6"	
		/clone="DAS-75011"	
		/clone_1lb="DNA-arcs-BAC.1-SSTO.1"	
		1..31974	
		/note="assembly fragment:00419	
		fragment_chain:1"	
		32075..51357	
		/note="assembly fragment:00594	
		fragment_chain:1"	
		51458..83889	
		/note="assembly fragment:00522	
		fragment_chain:1"	
		83990..94768	
		/note="assembly fragment:00308	
		fragment_chain:1"	
		-----	
		ORIGIN	
		Query Match 37.6%; Score 751.8; DB 14; Length 94768;	
		Best Local Similarity 81.9%; Pred. No. 4.6e-189;	
		Matches 944; Conservative 0; Mismatches 117; Indels 92; Gaps 3;	
		-----	
QY	1	AAAGGCTTGGAAATGACATGCGCTTCAACCTCTTATACCAACTCTGAGGTTGGG	60
Db	63182	AAAGGCTTGGAAATGACATGCGCTTCAACCTCTTATACCAACTCTGAGGTTGGG	63123
QY	61	GACATGCGCTTCCCTTCTAGGTCCTGTGACAGCAGCATCTGCTAATAGTCGACTTTGG	120
Db	63122	GACATGCGCTTCCCTTCTAGGTCCTGTGACAGCAGCATCTTACTAATAGTCGACTTTGG	63065
QY	121	GCCCTGATATTTTAACCTGTGTGCTCAAAATTTGTTCTCTAGAGTCAGGCGATCAAGCT	180
Db	63062	GCCCTGATATTTTAACCTGTGTGCTCAAAATTTGTTCTCTAGAGTCAGGCGATCAAGCT	63003
QY	181	ACAAGATGATCTTAACAAATGTAAACCCCAATGAGCTCACTAAACAATTTCTGCTGAGAGAC	240
Db	63002	ACAAGATGATCTTAACAAATGTAAACCCCAATGAGCTCACTAAACAATTTCTGCTGAGAGAC	62943
QY	241	CCTGGAACGACCGGCTGAGCCCTTTCATGAGCTTAAAGAGCTCCCTCTGAGAGCACTAC	300
Db	62942	CCTGGAATGATCACTAACCTTTTACTGAGCTTAAAGAGCTTCTCTGAGAGCACTAC	62888
QY	301	CACAGCAAGGAGCCCTTCTTCAACCCCTATCCAGAGAGAAATAGCTACAGCGGTATCGCA	360
Db	62882	AAATGCAAGGAGCCCTTCTTCAACCCCTATCCAGAGAGAAATAGCTATGATGATTTGCC	62822
QY	361	AAT-CCCAACGACGACTGGGCTGTCTCTGTTTGGAGGGGGGATTTGAGAGGTGAAGCCAGCT	419

Db 62822 AATCCCAACAGAGTTGCCATGTCCTGTTAGAGCGGAGATTGAGAGTCAAGT 62763  
420 GGGCTCTGGGTAGGAGGAGGAGTGGAGAACTTTTGTCTGCTAAAGATTGTAAT 479  
Db 62762 GGGCTCTGGGTAGGAGGAGGAGTGGAGAACTTTTGTCTGCTAAAGATTGTAAT 62703  
480 GCACCAATGAGCACTGTGTCTAGCTAAGAAATTGTAATGACCAATCAGACT 535  
Db 62702 GATCAATAGCACTGTGTCTAGCTAAGAAATTGTAATGACCAATCAGACT 62643  
536 ----- 535  
62642 AAAATGCAACCAATCAGTGTGTGTCTAGCTAAGAAATTGTAATGACCAATCACA 62583  
536 ----- CTGTAATGACCAATCAGAGATGTGGCGG 569  
62582 CTCTGTAATGACCAATCAGAGATGTGGCGG 62523  
570 GGTCAATTAAGAGGAGTAAAGAACTGGCAACCGGAGCAGAGTGGCAACCACTGGGTC 629  
62522 GGTCAATTAAGAGGAGTAAAGAACTGGCAACCGGAGCAGAGTGGCAACCTGGTCC 62463  
630 CCTTCACACTGTGAAAGCTTTTGTCTTCTTCACTAATTAATCTTGTCTGCTCA 689  
62462 ACTTCACAGCTGTGAAAGCTTTTGTCTTCTTCACTAATTAATCTTGTCTGCTCA 62403  
690 TTTCTTTGTCTTCACTAATTAATCTTGTCTTCTTCACTAATTAATCTTGTCTGCTCA 749  
62402 TTTCTTTGTCTTCACTAATTAATCTTGTCTTCTTCACTAATTAATCTTGTCTGCTCA 62343  
750 ATTCTGAAAGTCAAC-AGACCAAGAACCCAGTGAAGGAAAGAAAGTCCGAGTGTCT 808  
62342 ATTCTGAAAGGAGCAAGACCAAGAACCCAGTGAAGGAAAGAAAGTCCGAGTGTCT 62283  
809 GCTTTTAAGAGCTGTAACTCACTGAGAGTCTGAGCTTCACTCTGTAAGTCACTGA 868  
62282 ACCTTTAAGAGCTGTAACTCACTGAGAGTCTGAGCTTCACTCTGTAAGTCACTGA 62223  
869 GACCAACAAACCCAGAGGAAAGAACTGTGACACACTGAATTAATCTGAAAGAAACA 928  
62222 GACTAAGAAATCCACAGAGGAAAGAACTGTGACACACTGAATTAATCTGAAAGAAACA 62163  
929 CTCACACACACATCTTTCAGAGCTGTAACTCACTGAGAGTCTGAGCTTCACTCTGTA 988  
62162 CTCGACACACATCTTTCAGAGCTGTAACTCACTGAGAGTCTGAGCTTCACTCTGTA 62103  
989 TTGAAGTCAAGACCAAGAACCCAGGAGGAAAGAAATTCACAGACAGTGAAGAAAT 1048  
62102 TTGAAGTCAAGACCAAGAACCCAGGAGGAAAGAAATTCACAGACAGTGAAGAAAT 62043  
1049 CTGTATTTTGAAT 1061  
Db 62042 TCCCATTTTACAT 62030

RESULT 32  
BX927235 99242 bp DNA linear PRI 16-FEB-2005  
LOCUS Human DNA sequence from clone DAMA-25D9 on chromosome 6, complete  
DEFINITION sequence.  
ACCESSION BX927235  
VERSION BX927235.7 GI:59891245  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE 1 (bases 1 to 99242)  
AUTHORS Sycamore,N.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,

## COMMENT

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Feb 16, 2005 this sequence version replaced gi:45598013.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs constructed by  
the MHC Haplotype Consortium and collaborators. Further information  
can be found at  
<http://www.sanger.ac.uk/HGP/Chr6/MHC>  
DAMA-25D9 is from the DNA-Arts human BAC library MANN.1 VECTOR:  
pBelobAC11.

## FEATURES

## source

1. 99242  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone\_lib="DAMA-25D9"  
/clone\_lib="DNA-Arts BAC library MANN.1"

## ORIGIN

Query Match 37.6%; Score 751.8; DB 8; Length 99242;  
Best Local Similarity 81.9%; Pred. No. 4.6e-189;  
Matches 944; Conservative 0; Mismatches 117; Indels 92; Gaps 3;  
1 AAAAGCTTCTGAATGAGACATGCTTCAACTTTATACCACTCGAGTGGGC 60  
37525 AAAAGCTTCTGAATGAGACATGCTTCAACTTTATACCACTCGAGTGGGC 37584  
61 GACATGGCTTCCCTTTCTAGTCTGTGACAGCATCTGCTAATAGTGCATTTGG 120  
37585 GACATGGCTTCCCTTTCTAGTCTGTGACAGCATCTAATAGTGCATTTGG 37644  
121 GCCCTGATTTTAACTCTTGTGTAATTTGTTCTCTAGAGTGAAGCCATCAAGCT 180  
37645 GCCCTGATTTTAACTCTTGTGTAATTTGTTCTCTAGAGTGAAGCCATCAAGCT 37704  
181 ACAGATGATCTTAACCAATGTAACCCCAATGAGTCACTAACAATTTGTGAGAGACC 240  
37705 ACAGATGATCTTAACCAATGTAACCCCAATGAGTCACTAACAATTTGTGAGAGACC 37764  
241 CCGTGAACGACCGCTGCGCTTTCATAGTGGCTTAAAGAGCTCCCTGTGAGAGCACTAC 300  
37765 CCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 37824  
301 CACTGAGAGGCGCTTCTTCACTCTATTCAGAGAAATGATGATGATGATGATGATGATGATG 360  
37825 AACTGAAAGGCGCTTCTTCACTCTATTCAGAGAAATGATGATGATGATGATGATGATGATG 37884  
361 AAT-CCCAACAGAGCTGGGCTGTCTGTTTGAAGGGGAGTGAAGAGTGAAGCAGCT 419

[illegible]

COMMENT

Cambridgehire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Oct 12, 2004 this sequence request version replaced g1.5314569.  
----- Genome Center -----  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a VAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone comp15 constructed by the MHC Haplotype Consortium and collaborators. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrc/MHC> DMC-146617 is from the DNA-Arts.org BAC library MCF.1 VECTOR: pELOBAC11.

## FEATURES

**source**

```
1. .104114
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="DAMC-146E17"
/clone_lib="DNA-Arts.org"
```

ORIGIN

Query Match	37.6%	Score 751.8;	DB 8;	Length 104114;
Best Local Similarity	81.9%;	Pred. No. 4.6e-189;		
Matches 944; Conservative	0;	Mismatches 117;	Indels 92;	Gaps 3

1 AAAGCTTCTGAATCAGACCAATGCCCTTCAACTCTTATACCAACCTCTGAGTTGGGC 60

Db 74796 AAAGCTTCTGAATCAGACATGCTTCAAACTCTTATACCAACTCTGGAGTTGGC 74855

51 CACATGACATATCCACCTTCTACGCTCTTGACACGCCATCTTCTAATAGTCCGCAATTGG 120

100

Db 74856 GACATGGCTTCCTCCCTTCAGGTCCTGACAGCCATCTACAAATACCGCCCTTGG / 321

121 GCCCTGATTTTAACTCTGGTCAAAATTGTTTCCCTCTAGGATCGAGGCCATCAAGCT 180

Db 74916 GCCCTGAGTTTAACTCCTTGTCAAATTTGTTACTCTAGATCAAGGCCACAAGCT 74975

240

101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 109

Db 74976 ACAGATGCTTACAAAGGGAACCCCAATGAGCTCAGCTAACACACTTCTACCAAGGACC /5033

241 CCTGACCGACCGCTGGCCCTTCAATGGCCTAAGAGCTCCCTCTGGAGGACACTAC 300

75036 CCTGAAATGACTGACTGACCTTTTACTGGCCCTAAAGAGCTTCTCTCTGGAGGACACTAC 75095

[illegible]

301 CACIGCAGGCCCCIICACCCCAICCGACGGAAATGCTTCACCCCATTTCC

Db 75096 AACTGCAAGGCCCTTCTTCACCCCTATCCAGCAGAGTAGCTAGAGTGGTCATTGCC 75155

361 AAT-CCCAACAGCAGCTGGGGTGTCTGTTTGAGGGGGATTGAGAGGTGAAGCCAGCT 419

```

Db      75156 AATTCACACAGAGAGTTCATGCTCCGTTTAAAGCGGGGATTTGAGAGTGAGTCAAGCT 75215
QY      420 GGGCTCTGGGTGAGGGGGGAGCTTGGAGAACTTTTGTCTCTGCTTAAAGATTTGTAAT 479
Db      75216 GGGCTCTGGGTGAGGGGGGAGCTTGGAGAACTTTTGTCTCTGCTTAAAGATTTGTAAT 75275
QY      480 GCACCAATGAGCACTGTGTCTAGCTAAAGATTTGTAATGCAACCAATGAGCACT 535
Db      75276 GCATCAATGAGCACTGTGTCTAGCTAAAGATTTGTAATGCAACCAATGAGCACTGTGT 75335
QY      536 ----- 535
Db      75336 AAAATGACCAATCAGTGTCTGTGTCTAGCTAAATGATTTGTAATGACCAATCAACA 75395
QY      536 -----CTGTAATAATGACCAATCAGAGATGTGGCGG 569
Db      75396 CTCTGTAAATGAGCAATCAGACGCTGTAAATGACTTAATCAGTAAAGATCGGGCAG 75455
QY      570 GGTCAATAAGGAGTAAATACTGGCCAGCCGAGCCAGAGTGGCAACCCACTCGGGTCC 629
Db      75456 GGGCAATAAGGAGTAAATACTGGCCAGCCAGCCGAGCCAGAGTGGCAACCTGCTAGGTCC 75515
QY      630 CCTTCACACTGTGAGAGCTTTGTCTTTTGTCTTTTCACTTAATAATCTTGCTGCTCA 689
Db      75516 ACTTCACGCTGTGAGAGCTTTGTCTTTTCACTTTCACTTAATAATCTTGCTGCTCA 75575
QY      690 TTTCTTTGTCTCCACTACTCTTTATGAGCTTAACTCACTGCGAGGGTCTGTGGCTTC 749
Db      75576 CTCTTTGGGTTCAGACTACTCTTTATGAGCTTAACTCACTGCGAGGGTCTGTGGCTTC 75635
QY      750 ATTCTGGAAGTCAAC-AGACCAAGAACCCCACTGGAAGGAACAAGAACTCCCGATGTGCT 808
Db      75636 ATTCTGGAAGGCAAGACAGCAAGAACCACTCGGAGGAACAAGAACTCCCGATGTGCT 75695
QY      809 GCCTTTAAGAGCTTAACTCACTGCGAGGCTCTGAGCTTCACTCTGTAAGTCACTGA 868
Db      75696 ACCTTTAAGAGCTTAACTCACTGCGAGGCTCTGAGCTTCACTCTGTAAGTCACTGA 75755
QY      869 GACACCAAAACCCACCAAGGAAGAACTCTGGAACAACCTGTAATTTGGAAGGAACAA 928
Db      75756 GACTAGGAATCCACCAAGGAAGAACTCTGGAACAACCTGTAATTTGGAAGGAACAA 75815
QY      929 CTCGACACACCATCTTTTCAAGGCTGTAACTCACTCCGAGAGGCTGTGGCTTCAATTC 988
Db      75816 CTCGACACACCATCTTTTCAAGGCTGTAACTCACTCCGAGAGGCTGTGGCTTCAATTC 75875
QY      989 TTGAAGTCAAGCAAGCAACCAACCCAGGAAGGAACAAATTCACAGACAGTGAAGAT 1048
Db      75876 TTGAAGTCAAGTCAAGCAACCAACCCAGGAAGGAATTAATTCGACACAGAGAGCTG 75935
QY      1049 CTGTATTTTGTAT 1061
Db      75936 TCCCATTTTACAT 75948

```

```

RESULT 34
BX120007 109682 bp DNA linear PRI 18-MAY-2005
LOCUS Human DNA sequence from clone DASS-397D15 on chromosome 6, complete
DEFINITION BX120007
ACCESSION BX120007
VERSION 9
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE
1 (bases 1 to 109682)
AUTHORS Sycamore,N.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

```

## COMMENT

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
 Clone requests: clonerequest@sanger.ac.uk  
 On Jun 2, 2003 this sequence version replaced gj:30024440.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs constructed by the MHC Haplotype Consortium and collaborators. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6/MHC  
 DASS-397D15 is from a DNA-arts SSTO human bac library VECTOR: pBelOBAC11

Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

## source

1..109682

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="DASS-397D15"

/clone\_1fb="DNA-arts-BAC.1-SSTO.1"

/location/Qualifiers

1

/note="Clone\_left\_end: DASS-397D15"

12660

/note="Clone\_right\_end: DASS-380D18"

13315..13395

/note="Single clone region. Sequence generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."

56811..56841

/note="Sequence from uni-directional dGTP big dye terminator reads only."

109682

/note="Clone\_right\_end: DASS-397D15"

ORIGIN

Query Match 37.6%; Score 751.8; DB 8; Length 109682;

Best Local Similarity 81.9%; Pred. No. 4.6e-189;

Matches 944; Conservative 0; Mismatches 117; Indels 92; Gaps 3;

1

AAAGGCTCTGAATAGACATGCTTCAACTCTTATACCAACCTGAGTTGGGC 60

87451 AAAGGCTCTGAATAGACATGCTTCAACTCTTATACCAACCTGAGTTGGGC 87510

61 GACATGAGCTTCCCTCTTCTAGTCTGTCAGAGCATCTGTAATAGTCATTTGG 120

87511 GACATGAGCTTCCCTCTTCTAGTCTGTCAGAGCATCTGTAATAGTCATTTGG 87570

121 GCCCTGATTTTAACTCTTGTCAATTTGTTCTCTAGAGTCAGGCCATCAAGCT 180

87571 GCCCTGATTTTAACTCTTGTCAATTTGTTCTCTAGAGTCAGGCCATCAAGCT 87630

181 ACAGATGATCTTAACAATTAACCCCAATGAGCTCAACTCACTTGTGTCAGAGCC 240

87631 ACAGATGATCTTAACAATTAACCCCAATGAGCTCAACTCACTTGTGTCAGAGCC 87690

241 CCTGACGACCGCTGGCTTTGCAATGCTTAAAGAGCTCCCTCTGAGAGACACTAC 300

87691 CCTGACGACCGCTGGCTTTGCAATGCTTAAAGAGCTCCCTCTGAGAGACACTAC 87750

[illegible]

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Hominidae; Homo.  
1 (bases 1 to 197067)  
Sims, S.  
Direct Submission  
Submitted (30-JUL-2003) Wellcome Trust Sanger Institute, Hinxton  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
hmmguy@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jul 31, 2003 this sequence version replaced gi:32997038.

```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bSS105012
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Chemistry: Dye-terminator, 100% of reads
Consensus quality: 187436 bases at least Q40
Consensus quality: 190482 bases at least Q30
Consensus quality: 192676 bases at least Q20
Insert size: 194867, sum-of-contigs
Insert size: 163251, 16.1% error, agarose-fp
Quality coverage: 3.71x in Q20 bases; sum-of-contigs
coverage: 5.55x in Q20 bases; agarose-fp

```

\*-----\*

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 23 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

	* de preserved	17025:	contig of 17025 bp in length
*	1	17125:	gap of 100 bp
*	17026:	contig of 2913 bp in length	
*	17126	20038:	gap of 100 bp
*	20039	30881:	contig of 10743 bp in length
*	30882	30981:	gap of 100 bp
*	30982	41127:	contig of 10746 bp in length
*	411728	41827:	gap of 100 bp
*	41828	47939:	contig of 6012 bp in length
*	47840	47939:	gap of 100 bp
*	47940	62377:	contig of 14438 bp in length
*	62378	62477:	gap of 100 bp
*	62478	65276:	contig of 2799 bp in length
*	65277	65376:	gap of 100 bp
*	65377	72991:	contig of 7615 bp in length
*	72992	73091:	gap of 100 bp
*	73092	79639:	contig of 6448 bp in length
*	79540	79639:	gap of 100 bp
*	79640	85515:	contig of 5976 bp in length
*	85616	85715:	gap of 100 bp
*	85716	98517:	contig of 12802 bp in length
*	98518	98517:	gap of 100 bp
*	98618	103829:	contig of 5212 bp in length
*	103830	103829:	gap of 100 bp
*	103930	119341:	contig of 15412 bp in length
*	119342	119441:	gap of 100 bp
*	119442	127313:	contig of 7872 bp in length
*	127313	127413:	gap of 100 bp
*	127414	145480:	contig of 18077 bp in length
*	145491	145500:	gap of 100 bp
*	145591	155206:	contig of 9616 bp in length
*	155207	155306:	gap of 100 bp
*	155307	157763:	contig of 2457 bp in length
*	157764	157863:	gap of 100 bp
*	157864	162040:	contig of 4177 bp in length
*	162041	162140:	gap of 100 bp
*	162141	168837:	contig of 6697 bp in length
*	168838	168937:	gap of 100 bp
*	168938	180653:	contig of 11716 bp in length



\* 180654 180753: gap of 100 bp  
\* 180754 187735: contig of 6982 bp in length  
\* 187736 187835: gap of 100 bp  
\* 187836 194764: contig of 6929 bp in length  
\* 194765 194864: gap of 100 bp  
\* 194865 197067: contig of 2203 bp in length.  
Location/Qualifiers  
1. 197067

## FEATURES

## Source

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="DAS-105012"  
/clone\_1kb="DNA-arte-BAC.1-SSTO.1"  
1. 17025  
/note="assembly\_fragment:00433  
fragment\_chain:1"  
17126. 20038  
/note="assembly\_fragment:00355  
fragment\_chain:1"  
20139. 30881  
/note="assembly\_fragment:00385  
fragment\_chain:1"  
30982. 41727  
/note="assembly\_fragment:00872  
fragment\_chain:1"  
41828. 47839  
/note="assembly\_fragment:00083  
fragment\_chain:2"  
47940. 62377  
/note="assembly\_fragment:01078  
fragment\_chain:2"  
62478. 65276  
/note="assembly\_fragment:00283  
fragment\_chain:3"  
65377. 72991  
/note="assembly\_fragment:00572  
fragment\_chain:3"  
73092. 79539  
/note="assembly\_fragment:00648  
fragment\_chain:4"  
79640. 85615  
/note="assembly\_fragment:00488  
fragment\_chain:4"  
85716. 98517  
/note="assembly\_fragment:00715  
fragment\_chain:5"  
98618. 103829  
/note="assembly\_fragment:01139  
fragment\_chain:5"  
103930. 119341  
/note="assembly\_fragment:00046.0"  
119442. 127313  
/note="assembly\_fragment:00049"  
127414. 145490  
/note="assembly\_fragment:00276"  
145591. 155206  
/note="assembly\_fragment:00381.0"  
155307. 157763  
/note="assembly\_fragment:00593"  
157864. 162040  
/note="assembly\_fragment:00664"  
162141. 168837  
/note="assembly\_fragment:00713.0"  
168938. 180653  
/note="assembly\_fragment:00882"  
180754. 187735  
/note="assembly\_fragment:00940"  
187836. 194764  
/note="assembly\_fragment:01037"  
194865. 197067  
/note="assembly\_fragment:01039"

ORIGIN

Query Match 37.6%; Score 751.8; DB 14; Length 197067;  
Best Local Similarity 81.9%; Pred. No. 4.6e-189;  
Matches 944; Conservative 0; Mismatches 117; Indels 92; Gaps 3;

QY 1 AAAGCTTGAATGACCAATGACCTTCAACTTTATACCACTCTGAGTTGGG 60  
DB 11498 AAAGCTTGAATGACCAATGACCTTCAACTTTATACCACTCTGAGTTGGG 11439  
QY 61 GACATGGCTTCTCCCTTTCTAGGCTCTGTGACAGCACTTGTATATGCGATTGG 120  
DB 11438 GACATGGCTTCTCCCTTTCTAGGCTCTGTGACAGCACTTGTATATGCGATTGG 11379  
QY 121 GCCGTATATTTTAACTCTGTGGTCAATTTGTTCTCTAGATGAGGCGCATCAAGCT 180  
DB 11378 GCCGTATATTTTAACTCTGTGGTCAATTTGTTCTCTAGATGAGGCGCATCAAGCT 11319  
QY 181 ACAGATGATCTTACCAATGTAACCCCAATGAGCTCACTAACCTTCTGTAGAGACC 240  
DB 11318 ACAGATGATCTTACCAATGTAACCCCAATGAGCTCACTAACCTTCTGTAGAGACC 11259  
QY 241 CCTGACCAACCCGCTGCGCTTTCAATGACCTCCCTCTGAGGACACTAC 300  
DB 11258 CCTGACCAACCCGCTGCGCTTTCAATGACCTCCCTCTGAGGACACTAC 11199  
QY 301 CACTGAGGCGCCCTTCTTCAACCCCTATCAGAGAAATGATACAGCGTATCGCCA 360  
DB 11198 CACTGAGGCGCCCTTCTTCAACCCCTATCAGAGAAATGATACAGCGTATCGCCA 11139  
QY 361 AAT-CCCAACAGACGTGGGGTGTCTGTTTGGAGGGGGATTGAGAGTGAAGCCAGCT 419  
DB 11138 AAT-CCCAACAGACGTGGGGTGTCTGTTTGGAGGGGGATTGAGAGTGAAGCCAGCT 11079  
QY 420 GGGCTTCTGGGTGAGGTGGGAACTTGGAACTTGTGTCTAGCTAAAGATTGTAAAT 479  
DB 11078 GGGCTTCTGGGTGAGGTGGGAACTTGGAACTTGTGTCTAGCTAAAGATTGTAAAT 11019  
QY 480 GCACCAATGACACTGTGTGTCTAGCTAAAGATTGTAAATGACCAATCAGACT- 535  
DB 11018 GCACCAATGACACTGTGTGTCTAGCTAAAGATTGTAAATGACCAATCAGACTGTGT 10959  
QY 536 ----- 535  
DB 10958 AAAAATGACCAATCAGTGTCTGTGTCTAGCTAAATGATTGTAATGACCAATCAACA 10899  
QY 536 -----CTGTAAATGACCAATCAGAGATGTGGCG 569  
DB 10898 CTGTGTAAATGACCAATCAGAGATGTGGCG 10839  
QY 570 GGTCAATTAAGGAGTAAATCTGGCAACCCGAGCGAGCAAGTGGCAACCACTCGGGTGC 629  
DB 10838 GGTCAATTAAGGAGTAAATCTGGCAACCCGAGCGAGCAAGTGGCAACCACTCGGGTGC 10779  
QY 630 CCTTCACACTGTGAAAGCTTTGTTTGTCTTACCAATTAATCTGTGCTGCTCA 689  
DB 10778 ACTTCACACTGTGAAAGCTTTGTTTGTCTTACCAATTAATCTGTGCTGCTCA 10719  
QY 690 TTCTTGTGTCACTACTTATGAGCTGTAACTCACTGCGAGGCTGTGTGCTTC 749  
DB 10718 TTCTTGTGTGTCACTACTTATGAGCTGTAACTCACTGCGAGGCTGTGTGCTTC 10659  
QY 750 ATTCTGAAGTGAAC-AGACACGAACCACTGGAAGGAACAAGAACTCCGATGTCT 808  
DB 10658 ATTCTGAAGTGAAGCAAGCAACCACTGGAAGGAACAAGAACTCCGATGTCT 10599  
QY 809 GCTTTAAGAGCTGTAACTCACTGCGAAGCTTCACTCTGGAAGTCACTGA 868  
DB 10598 GCTTTAAGAGCTGTAACTCACTGCGAAGCTTCACTCTGGAAGTCACTGA 10539  
QY 869 GACCAAAACCAACGAAGAAAGAACTTGAACCACTTGAATTTGAAAGAAACA 928  
DB 10538 GACCAAAACCAACGAAGAAAGAACTTGAACCACTTGAATTTGAAAGAAACA 10479



QY 480 ----GCACCAATGACGACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACT 535  
| | | | |  
Db 64972 AAGGACCAATGACGACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACT 64913  
| | | | |  
QY 536 CTGT-AAAATGACCAATGACGAGATGTGGCGGGGTCAATTAAGGAGTAAATGCG 594  
| | | | |  
Db 64912 CTGTAAAAATGACCAATGACGAGATGTGGGTGGGGCAAAATAGGAATTAAGAGTAG 64853  
| | | | |  
QY 595 CCACCCGAGCAGACAGTGGGCAACCCACTGGGGTCCCTTCCA-CACGTGTGAAGCTTTGT 653  
| | | | |  
Db 64852 CCACCCGAGCAGCAGACAGCAACAACCTGGGGTCCCTTCCAACACTGTGGAAGCTTTGT 64793  
| | | | |  
QY 654 TCTTTGCTCTCTCAATAAATCTGTGTCTGCTCATTTTGTGTGTCACACTACCTTTA 713  
| | | | |  
Db 64792 TCTTCACTCTTGTCAATTAATCTGTGTCTGCTCATTTTGGGTGTGACCTATCTTTA 64733  
| | | | |  
QY 714 TGAAGCTGTAACTCACTGCGAGGGTCTGTGGCTTCAATTCCTGAAGTCA-ACAGACGAG 772  
| | | | |  
Db 64732 TGAAGCTGTAACTCACTGCGAGGGTCTGTGGCTTCACTCTGAAGTCAATGACGAG 64673  
| | | | |  
QY 773 AACCCACTGGAAGGAACAAAGAACTCCCATGTGTGCTCTTTAAGCTGTAACTACTAC 832  
| | | | |  
Db 64672 AACCCACTGGAAGGAACAACTCCGACGGCCAACTTAAGAGCTGTAACTACTAC 64613  
| | | | |  
QY 833 TGGGAAGCTCTGAGCTTCACTCCTGAAGTCAAGTGAACCAACCAACCAAGGAAG 892  
| | | | |  
Db 64612 TGGGAAGCTCTGAGCTTCACTCCTGAAGTCAAGTGAACCAACCAACCAAGGAAG 64553  
| | | | |  
QY 893 AAACCTGTGACACACCTGTAATCTGAAGGAACAACTCCAGACACACATCTTTGAGAG 952  
| | | | |  
Db 64552 AAACCTGTGACACATCTGAAGTCAAGTGAACCACTGACACACCACTTTTAAGAA 64493  
| | | | |  
QY 953 CTGTAACTCACTACCCGAGGGTCTGTGGCTTCAATTTCTGAAGTCAAGGAACCAAGAAC 1012  
| | | | |  
Db 64492 CTGTAACTCACTACCCGAGGGTCTGTGGCTTCAATTTCTGAAGTCAAGGAACCAAGAAC 64433  
| | | | |  
QY 1013 CACCGAAGGAACAAATTCAGACAGATGAAGAACT 1050  
| | | | |  
Db 64432 CACCGAAGGAACCAATTCAGACAGAGGATCACT 64395  
| | | | |  
RESULT 37  
AC145142 184449 bp DNA linear HTG 10-JUN-2003  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-1398P22, WORKING DRAFT  
SEQUENCE, 3 unordered pieces.  
ACCESSION AC145142  
VERSION AC145142.1 GI:31560141  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 184449)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 184449)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUN-2003) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 2756848  
Center clone name: RPCI-11\_1398P22  
-----  
Summary Statistics

Consensus quality: 184227 bases at least Q40  
Consensus quality: 184248 bases at least Q30  
Consensus quality: 184249 bases at least Q20  
Estimated insert size: 175000; agarose-fp estimation  
Quality coverage: 12.87 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 12.22 in Q20 bases; agarose-fp estimation  
Quality coverage: 12.22 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1188: contig of 1188 bp in length  
\* 1189 1288: gap of unknown length  
\* 1289 63133: contig of 61845 bp in length  
\* 63134 63233: gap of unknown length  
\* 63234 184449: contig of 121216 bp in length.  
FEATURES  
Location/Qualifiers  
source  
1. 184449  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone\_lib="RP11-1398P22"  
/clone\_1lb="RPCI human BAC library 11"  
gap 1189. 1288  
gap /estimated\_length=unknown  
gap 63134. 63233  
/estimated\_length=unknown  
ORIGIN  
Query Match 37.3%; Score 745.6; DB 14; Length 184449;  
Best Local Similarity 84.1%; Pred. No. 26-187;  
Matches 940; Conservative 0; Mismatches 109; Indels 69; Gaps 6;  
QY 1 AAAGCTTGTGAATGACAAATGCTTTCAAACTTTATACCACTCTGAGTTGGG 60  
| | | | |  
Db 181095 AAAGCTTGTGAATGACAAATGCTTTCAAACTTTATACCACTCTGAGTTGGG 181154  
| | | | |  
QY 61 GACATGGCTTTCCTCCCTTTCTAGTCTGTGACAGCACTGCTAATGCTGCATTGG 120  
| | | | |  
Db 181155 AACATGGCTTTCCTCCCTTTCTAGTCTGTGACAGCACTGCTAATGCTGCATTGG 181214  
| | | | |  
QY 121 GCCGTGTATTTTAACTCTGTGTCAAATTTTCTCTAGAGATGAGGCACTAAGCT 180  
| | | | |  
Db 181215 GCCGTGTATTTTAACTCTGTGTCAAATTTTCTCTAGAGATGAGGCACTAAGCT 181273  
| | | | |  
QY 181 ACAGATGATCTTAAATGTAACCCCAATGAGCTTCACTAACACTTGTGTGAGACC 240  
| | | | |  
Db 181274 ACAGATGATCTTAAATGTAACCCCAATGAGCTTCACTAACACTTGTGTGAGACC 181333  
| | | | |  
QY 241 CCTGACACGACCCCGCTGCTTTTCAATGAGCTTCCCTCTGAGAGCACTAC 300  
| | | | |  
Db 181334 CCTGACACGACCCCGCTGCTTTTCAATGAGCTTCCCTCTGAGAGCACTAC 181393  
| | | | |  
QY 301 CACTGAGGAGCCCTTCTTCACTCCCTATCCAGAGAGTACTACAGGCTCATGCGCA 360  
| | | | |  
Db 181394 CACTGAGGAGCCCTTCTTCTTCTTCACTCCCTATCCAGAGAGTACTACAGGCTCATGCGCA 181453  
| | | | |  
QY 361 AA-TCCCAACAGACAGTGGGTGTCTGTTTGAAGGGGAGATTGAGGTGAAGCAGCT 419  
| | | | |  
Db 181454 AATTCACCAACAGAGTGGGTGTCTGTTTGAAGGGGAGATTGAGGTGAAGCAGCT 181513  
| | | | |  
QY 420 GGGCTTCTGGGTCAAGTGGGAGCTTGAAGAACTTTTGTCTACGCTAAAGATTTGAAT 479  
| | | | |  
Db 181514 GGGCTTCTGGGTCAAGTGGGAGCTTGAAGAACTTTTGTCTACGCTAAAGATTTGAAT 181573  
| | | | |  
QY 480 ----- 479  
| | | | |  
Db 181574 GCACCAATGACGACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACTCTGTAA 181633  
| | | | |

QY 480 -----GCACCAATCAGCACTCTGTCTAGCTAAAGGATTGTAATGACCAATCAGCACT 535  
Db 181634 AAACGACCAATCAGCACTCTGTCTAGCTAAAGGATTGTAATGACCAATCAGCACT 181693  
QY 536 CTGT-AAAATGAGCAATCAGCAGGATGTGGCGGGGTCAAAATTAAGGAGTAAAACTGG 594  
Db 181694 CTGTAAAATGAGCAATCAGCAGGATGTGGCGGGGTCAAAATTAAGGAGTAAAACTGG 181753  
QY 595 CCACCCGAGCAGCAGTGGCAACCACTCGGGTCCCTTCCCA-CACTGTGGAGCTTTGT 653  
Db 181754 CCACCCGAGCAGCAGCAGCAGCACTCGGGTCCCTTCCCACTGTGGAGCTTTGT 181813  
QY 654 TCTTTTGTCTTCACAATTAATCTTGTCTGTCTGCTATCTTTGTGTGCACTACCTTTA 713  
Db 181814 TCTTTTGTCTTCACAATTAATCTTGTCTGTCTGCTATCTTTGTGTGCTGCACTATCTTTA 181873  
QY 714 TGAGCTGTAACTACTCTGAGGCTGTGTGCTTCAATTCCTGAAGTCA-ACAGACCAG 772  
Db 181874 TGAGCTGTAACTACTCTGAGGCTGTGTGCTTCAATTCCTGAAGTCA-ACAGACCAG 181933  
QY 773 AACCCAGTGAAGAAACAAGAACTCCGATGTGTCTCTTTAAGCTGTAACTCTAC 832  
Db 181934 AACCCAGTGAAGAAACAAGAACTCCGATGTGTCTCTTTAAGCTGTAACTCTAC 181993  
QY 833 TCGGAAGCTCTGAGCTTCACTCTGAGTCAAGTCAAGACCAACCAACCAAGAGAG 892  
Db 181994 TCGGAAGCTCTGAGCTTCACTCTGAGTCAAGTCAAGACCAACCAACCAAGAGAG 182053  
QY 893 AAACCTGAGACACCTGATATCTGAAGAGAACTCCAGACAGACCACTCTTCAAG 952  
Db 182054 AAACCTGAGACACCTGATATCTGAAGAGAACTCCAGACAGACCACTCTTCAAG 182113  
QY 953 CTGTAACTACTCAGCGAGGCTGTGTCTTCAATCTTGAAGTCAAGACCAAGAGAG 1012  
Db 182114 CTGTAACTACTCAGCGAGGCTGTGTCTTCAATCTTGAAGTCAAGAGAGAGAGAG 182173  
QY 1013 CACCGGAAGAAACAATCTCAGACACAGTGAAGAAATCT 1050  
Db 182174 CACCGGAAGAAACAATCTCAGACACAGTGAAGAAATCT 182211

RESULT 38  
AC074133/c 75852 bp DNA linear PRI 28-FEB-2002  
LOCUS AC074133  
DEFINITION Homo sapiens chromosome 5 clone CTD-2278B20, complete sequence.  
ACCESSION AC074133  
VERSION AC074133.5 GI:18997252  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE  
AUTHORS 1. (bases 1 to 75852)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
REFERENCE 2. (bases 1 to 75852)  
AUTHORS DOE Joint Genome Institute.  
JOURNAL Direct Submission  
REFERENCE 3. (bases 1 to 75852)  
AUTHORS Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint  
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
TITLE Direct Submission  
AUTHORS Submitted (23-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell  
JOURNAL Drive, Walnut Creek, CA 94598, USA  
REFERENCE 4. (bases 1 to 75852)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
TITLE Submitted (28-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell  
JOURNAL Drive, Walnut Creek, CA 94598, USA

COMMENT On Feb 28, 2002 this sequence version replaced gi:18266630.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99% of Sequence;  
Estimated Total Number of Errors is 0.6.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 112kb). It is clipped at the overlap with AC008883. The  
number of bases overlapped is 13829.

FEATURES  
source  
location/Qualifiers  
1..75852  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2278B20"

ORIGIN  
Query Match 37.1%; Score 742.2; DB 8; Length 75852;  
Best Local Similarity 85.6%; Pred. No. 1.7e-186;  
Matches 936; Conservative 0; Mismatches 108; Indels 49; Gaps 8;

QY 1 AAAGCTTGTGAATCAGACAAATGCTTCAACTATACCACTCTGGAGTTGGG 60  
Db 36812 AAAGCTTGTGAATCAGACAAAGCTTCAACTATACCACTCTGGAGTTGGG 36753  
QY 61 GACATGGCTTCCCTCTTCTAGTCTGTGACAGCACTCTGTAATAGTCATTTGG 120  
Db 36752 AACCTGGCTTCCCTCTTCTAGTCTGTGACAGCACTCTGTAATAGTCATTTGG 36693  
QY 121 GCCCTGATATTTTAACTCTTGTGTCAAATTTGTTCTCTAGATGAGGCATAGCT 180  
Db 36692 GCCCTGATATTTTAACTCTTGTGTCAAATTTGTTCTCTAGATGAGGCATAGCT 36634  
QY 181 ACAGATGATCTTCAAAATGTAACCCCAATAGCTCACTAACTTCTGTGAGGAGC 240  
Db 36633 ACAGATGATCTTCAAAATGTAACCCCAATAGCTCACTAACTTCTGTGAGGAGC 36574  
QY 241 CCTGAGCGAGCCGCTGCTTCAATGAGCCCTTAAGAGCTCCCTCTGAGGAGACTAG 300  
Db 36573 CCTGAGCGAGCCGCTGCTTCAATGAGCCCTTAAGAGCTCCCTCTGAGGAGACTAG 36514  
QY 301 CACTGAGGAGCCCTCTTCACTCCCTATCCAGAGGAGTACAGCGGTCA-TCGCC 359  
Db 36513 AATTGTAAGGAGCCCTCTTCACTCCCTATCCAGAGGAGTACAGCGGTCA-TCGCC 36455  
QY 360 AAATCCCAACAGACGCTGGGGGTCTCTGTTTGAAGGGGGGATTGAAGTGAAGCAAGCT 419  
Db 36454 AAATCCCAACAGACGCTGGGGGTCTCTGTTTGAAGGGGGGATTGAAGTGAAGCAAGCT 36395  
QY 420 GGGCTTCTGGGTCAGGTCGGGGAATTTGAGAACTTT----- 454  
Db 36394 GGGCTTCTGGGTCAGGTCGGGGAATTTGAGAACTTTCTGTCTAGTAAAGATTGTAAGC 36335  
QY 455 -----TGTCCTAGCTAAAGATTGTAATGCAATCAGCACTCTGT 498  
Db 36334 ACAACATCAGCCCTCTGTGTCTAGCTAAAGATTGTAATGCAATCAGCACTCTGT 36275  
QY 499 GTCTAGCTAAAGATTGTAATGCAATCAGCACTCTGTAAATAGCAATTCAGCAG 558  
Db 36274 GTCTAGCTAAAGATTGTAATGCAATCAGCACTCTGTAAATAGCAATTCAGCAG 36215  
QY 559 GATGTGGCGGGGTCAATTAAGGAGTAAAGTGGCAACCGAGCAGCAGTGGCAAC 618  
Db 36214 GATGTGGG-TGAGCGCAATTAAGGAGTAAAGTGGCAACCGAGCAGCAGTGGCAAC 36156  
QY 619 CACTCGGATCCCTTCCACACTGTGAAGCTTTGTTCTTGTGCTTCAATTAATCTT 678  
Db 36155 TACTCAGGTCCCTTCCACACTGTGAAGCTTTGTTCTTGTGCTTCAATTAATCTT 36096  
QY 679 GCTGCTGCTCAATTTTGTGTCAACTACTTATAGCT--GTAACTCACTGCGAG 736

Db 36095 GCTGCTGCTCATTCTTTGGGTGACGACTCTTTATGAGTACATTAACACTCTGTGA 36036  
Qy 737 GGTCTGTGCTTCTTCTTGAAGTCAAC-AGACCAGCAACCCCTGGAAGCAAAAGAA 795  
Db 36035 GGTCTGTGCTTCTTCTTGAAGTCAAC-AGACCAGCAACCCCTGGAAGCAAAAGAA 35976  
Qy 796 CTCCTGAGTGTGCTTCTTGAAGTCAACACTCTGGAAGCTTGTGAGCTTCACTC 855  
Db 35975 CTCCTGAGTGTGCTTCTTGAAGTCAACACTCTGGAAGCTTGTGAGCTTCACTC 35916  
Qy 856 CTGAAGTCT-AGTAGACCAACAAACCCAGCAAGAAAGAACTGTGACACACTGAA 914  
Db 35915 CTGAAGTCTGAAGCAAGCAAGCAACCCAGCAAGAAAGAAAGAACTGTGACACACTGAA 35856  
Qy 915 TCTGAAGCAACAACTCTGACACACCATCTTTCAGAGCTGTAACACTGACCCAGAGGT 974  
Db 35855 TCTGAAGCAACAACTCTGACACACCATCTTTCAGAGCTGTAACACTGACCCAGAGGT 35796  
Qy 975 CTGTGCTTCTTCTTGAAGTCAAGCAAGCAAGCAACCCAGCAAGCAAAATTCAG 1034  
Db 35795 CCGTGTCTTCTTCTTGAAGTCAAGCAAGCAAGCAACCCAGCAAGCAAAATTCAG 35736  
Qy 1035 ACACAGTAGGAAA 1047  
Db 35735 ACACAGTAGGAAA 35723

## RESULT 39

AC091977/c

LOCUS AC091977 183494 bp DNA linear PRI 15-DEC-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-528L24, complete sequence.  
AC091977  
VERSION AC091977.3 GI:17861016

## KEYWORDS

HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.

## REFERENCE

1 (bases 1 to 183494)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 183494)

## REFERENCE

DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 183494)

## REFERENCE

DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (15-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

## REFERENCE

On Dec 15, 2001 this sequence version replaced gi:15290435.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov

## COMMENT

Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.7% of sequence;  
Estimated Total Number of Errors is 0.3.

## FEATURES

source

Location/Qualifiers  
1..183494  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-528L24"

## ORIGIN

Query Match 37.1%; Score 742.2; DB 8; Length 183494;  
Best Local Similarity 83.8%; Prid. No. 1.6e-186;  
Matches 918; Conservative 0; Mismatches 128; Indels 49; Gaps 5;

Qy 1 AAAGCTTGAATATGACAAATGCTTTCAAACTTATACCAACTCTGAGTTGGGC 60  
Db 35692 AAAGCTTGAATATGACAAATGCTTTCAAACTTATACCAACTCTGAGTTGGGT 35633  
Qy 61 GACATGGCTTCTCCCTTTCTAGAGTCTGTGACAGCATTTGTATATGTCGATTGG 120  
Db 35632 GACATGGCTTCTCTTCTAGAGTCTGTGACAGCATTTGTATATGTCGATTGG 35573  
Qy 121 GCCCTGATTTTAACTCTGTGCTAAATTTGTTTCTCTAGAGTGAAGGCCATCAAGCT 180  
Db 35572 GCCCTGATTTTAACTCTGTGCTAAATTTGTTTCTCTAGAGTGAAGGCCATCAAGCT 35513  
Qy 181 ACAGATGATCTTCAATATGTAACCCCAATAGCTCAATCAACTTCTGTGAGGACC 240  
Db 35512 ACAGATGATCTTCAATATGTAACCCCAATAGCTC-----AACAATTCTACGAGATC 35457  
Qy 241 CTGGAACGACCCGCTGCTCTTCAATGAGCTTAAAGCTCCCTCTGTGAGCACTAC 300  
Db 35456 CTGGAACGACCCGCTGCTCTTCAATGAGCTTAAAGCTCCCTCTGTGAGCACTAC 35397  
Qy 301 CACTGAGAGGCCCCCTTCTTCAACCCCTATCAGAGAGAGTACAGCGGTATC-GCC 359  
Db 35396 AACTGAGAGGCCCCCTTCTTCAACCCCTATCAGAGAGAGTACAGCGGTATCAACC 35337  
Qy 360 AATCCCAACAGACGCTGGGTGTCTCTGTTTGAAGGGGAGATTGAGAGTGAAGCCAGCT 419  
Db 35336 AATCCCAACAGACGCTGGGTGTCTCTGTTTGAAGGGGAGATTGAGAGTGAAGCCAGCT 35277  
Qy 420 GGGCTT-CTGGGTCAAGTGGGGAATTGAGAACTTTGTCTAGCTAAAGATTGTAA 478  
Db 35276 GGGCTTCTGGGTCAAGTGGGGAATTGAGAACTTTGTCTAGCTAAAGATTGTAA 35217  
Qy 479 TGACCAATCAG-----CACTCT 496  
Db 35216 TGACCAATCAG-----CACTCT 35157  
Qy 497 GTGTCTAGCTTAAAGATTGTAATGACCAATCAGACTCTGTAATGAGCAATCAGC 556  
Db 35156 GTGTCTAGCTTAAAGATTGTAATGACCAATCAGACTCTGTAATGAGCAATCAGC 35097  
Qy 557 AGGATGTGGGCGGGGTCAATATGAGGATTAAGGCTGGCCACCCGACCGACAGCTGAGCA 616  
Db 35096 AGGATGTGGGCGGGGTCAATATGAGGATTAAGGCTGGCCACCCGACCGACAGCTGAGCA 35037  
Qy 617 CCCACTGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTGTCTTCAATTAATC 676  
Db 35036 CCCACTGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTGTCTTCAATTAATC 34977  
Qy 677 TTGCTGTGCTGATTTGTTGTGCACTACCTTATAGAGCTGTAACTCACTGCGAG 736  
Db 34976 TTGCTGTGCTGATTTGTTGTGCACTACCTTATAGAGCTGTAACTCACTGCGAG 34917  
Qy 737 GGTCTGTGCTTCAATCTGAAGTCAAC-AGACCAAGAACCCACTGGAAGAAACAAGAA 795  
Db 34916 GGTCTGTGCTTCAATCTGAAGTCAAC-AGACCAAGAACCCACTGGAAGAAACAAGAA 34857  
Qy 796 CTCCTGAGTGTGCTTCTTGAAGTCAACACTCTGGAAGCTTGTGAGCTTCACTC 855  
Db 34856 CTCCTGAGTGTGCTTCTTGAAGTCAACACTCTGGAAGCTTGTGAGCTTCACTC 34797  
Qy 856 CTGAAGTCAAGTGAACCAACCCAGCAAGAAAGAAAGTGTGACCACTGAAATAT 915  
Db 34796 CTGAAGTCAAGTGAACCAACCCAGCAAGAAAGAAAGTGTGACCACTGAAATAT 34737  
Qy 916 CTGAAGCAACAACTCTGACACACCATCTTCAAGAGTGTAACTCACTGCAAGAGTCT 975  
Db 34736 CTGAAGCAACAACTCTGACACACCATCTTCAAGAGTGTAACTCACTGCAAGAGTCT 34677  
Qy 976 TGTGCTTCTTCTTGAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTTCGAA 1035  
Db 34676 TGTGCTTCTTCTTGAAGTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTTCGGA 34617  
Qy 1036 CACAGTAGGAATCT 1050

Db 34616 CACAATTAAACATGT 34602

|||||

RESULT 40 159797 bp DNA linear PRI 21-APR-2005  
AC104697

LOCUS Homo sapiens BAC clone RP11-454B3 from 2, complete sequence.  
DEFINITION AC104697 AC024403  
AC104697.2 GI:18855149  
VERSION  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE 1 (bases 1 to 159797)  
AUTHORS Cotton, M. and Radionenko, M.  
TITLE The sequence of Homo sapiens BAC clone RP11-454B3  
JOURNAL Unpublished (2001)  
REFERENCE 2 (bases 1 to 159797)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 159797)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
4 (bases 1 to 159797)  
AUTHORS Wilson, R.K.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Feb 21, 2002 this sequence version replaced gi:17933873.

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@watsn.wustl.edu  
----- Summary Statistics  
Center Project name: H\_NH0454B03  
Drafting Center: WIBR  
-----

NOTICE:  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:  
The RPC11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-337B18, 2000 bp overlap;  
the clone sequenced to the right is RP11-241F6, 2000 bp overlap.  
Actual start of this clone is at base position 167881 of  
RP11-337B18; actual end is at base position 2890 of RP11-241F6.  
Polymorphisms have been identified between AC092632, AC096560, and  
AC024403.

## FEATURES

The sequence of AC024403 has been incorporated into AC104697.

source Location/Qualifiers

1..159797  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-454B3"  
/clone\_11b="RPC1-11"  
misc\_feature 55931..56157  
/note="CPG island (%GC=62.6, o/e=0.91, #CPGs=20)"  
misc\_feature 59015..59288  
/note="Opg\_island (%GC=59.9, o/e=0.74, #CPGs=20)"

## ORIGIN

Query Match 37.0%; Score 741.2; DB 8; Length 159797;  
Best Local Similarity 85.1%; Pred. No. 3.1e-186;  
Matches 921; Conservative 0; Mismatches 128; Indels 33; Gaps 7;

1 AAGGCTTCTGAAATAGACATGCTTTCAAACTTTATACCACTCTGAGTTGGC 60  
29127 AAGGCTTCTGAAATAGACATGCTTTCAAACTTTATACCACTCTGAGTTGGC 29186  
61 GACATGCTTCCTCCCTTCTAGAGTCTGTGACAGCCATCTTCTATAGTGCATTTG 120  
29187 AACATGATTTTCCCTTTCTAGAGTCCATGACGACATCTTCTATAGTGCATTTG 29246  
121 GCCCTGATTTTAACTCTTGTGTCAAATTTGTTTCTCTAGAGTGCAGCCATCAAGCT 180  
29247 GCCCTGATTTTAACTCTTGTGTCAAATTTGTTTCTCTAGAGTGCAGCCATCAAGCT 29305  
181 ACAGATGATCTTCAAAATGTAACCCCAATGAGCTCACTAACAATCTTCTGAGAGCC 240  
29306 ACAGATGATCTTCAAAATGTAACCCCAATGAGCTCACTAACAATCTTCTGAGAGCC 29365  
241 CCGGACCGACCGCGCTGACCTTCAATGAGCTCAATGAGCTCCCTCTGAGAGCACTAC 300  
29366 CCGGACCGACCGCGCTGACCTTCAATGAGCTCAATGAGCTCCCTCTGAGAGCACTAC 29425  
301 CACTGACGAGCCCTTCTTCAACCCCTATCCAGCAGGAATGACTAGACGGTCAATC-GCC 359  
29426 AACTACAGAGTCCCTTCTTGTCCCTATCCAGCAGGAATGACTAGACGGTCAATC-GCC 29485  
360 AATTCACACAGACGCTGGGGTCTCTGTTGAGAGGGGATTTGAGAGTGAAGCCAGCT 419  
29486 AATTCACACAGACGCTGGGGTCTCTGTTGAGAGGGGATTTGAGAGTGAAGCCAGCT 29545  
420 GGGCTTCTGGGTGAGGTGGGAGCTTGGAGAACTTTGTGCTAGCTAAAGATTGTAAT 479  
29546 GGGCTTCTGGGTGAGGTGGGAGCTTGGAGAACTTTGTGCTAGCTAAAGATTGTAAT 29605  
480 GCACCAATCAGACCTCTGTGCTAGCTAAAGATTGTAATGACCAATCAGACCTCTGT 539  
29606 GCACCAATCAGACCTCTG-----TAAACACACCAATCAGACCTCTGT 29649  
540 AAATGACCAATCAGACGATGTGGCGGGGTCAATTAAGGAGTAAATCTGGCCACC 599  
29650 AAATGACCAATCAGACGATGTGGCGGGGTCAATTAAGGAGTAAATCTGGCCACC 29709  
600 CGAGCCAGCAGTGGCAACCACTCGGGTCCCTTCCACACTGAGGAAGCTTTGTTCTTT 659  
29710 CGAGCCAGCAGCAGCAACCTGCTGGGGTCCCTTCCACAC--TGAAGCTTTGTTCTTT 29767



QY 660 GCTCTTCAATAAATCTGCTGCTCAATCTTTGTGTCACACTACCTTATGAGCT 719  
Db 29768 GCTCTTCAATAAATCTGCTGCTCAATCTTTGTGTCACACTACCTTATGAGCT 29827  
QY 720 GTAACTACTGCGAGGCTGTGCTTATCTTAAAGTCAAC AGACCAAGAACCTA 778  
Db 29828 ATAACTACTGCGAGGCTGTGCTTATCTTAAAGTCAACAGAACCAAGAACCTA 29887  
QY 779 CTGGAAGAAACAAGAACTCCCGATGCTGCTTAAAGCTGTAACTCACTGCGAA 838  
Db 29888 CTGGAAGAAACA-----ATGACCACTTTAAAGCTTAACTCACTGCGAA 29936  
QY 839 GCTCTGAGCTTCACTCTGAGTCAAGTCAAGTCAAGAACCAAGAACCAAGT 897  
Db 29937 GGTCTGAGCTTCACTCTGAGTCAAGTCAAGTCAAGAACCAAGTCAAGT 29996  
QY 898 CTGGAACACCTGTAATCTGGAAGAAACAAGTCAAGTCAAGTCAAGTCAAGT 957  
Db 29997 CTGGAACACCTGTAATCTGGAAGAAACAAGTCAAGTCAAGTCAAGTCAAGT 30056  
QY 958 ACAGTCAAGCTGAGGCTGCTGCTTCAATCTTGAAGTCAAGTCAAGTCAAGTCAAGT 1017  
Db 30057 ACAGTCAAGCTGAGGCTGCTGCTTCAATCTTGAAGTCAAGTCAAGTCAAGTCAAGT 30116  
QY 1018 GAAGAAACAATCTGAGAACTGATTTTGAATCTGAGTCAAGTCAAGTCAAGT 1077  
Db 30117 GAAGAAACAATCTGAGAACTGATTTTGAATCTGAGTCAAGTCAAGTCAAGT 30176  
QY 1078 TA 1079  
Db 30177 TA 30178

RESULT 41  
AC034238/c 117583 bp DNA linear PRI 31-OCT-2001  
LOCUS AC034238 Homo sapiens chromosome 5 clone CTD-2313F11, complete sequence.  
DEFINITION AC034238  
AC034238  
VERSION AC034238.4 GI:16554348  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 117583)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
AUTHORS Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 117583)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
AUTHORS Submitted (31-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Oct 31, 2001 this sequence version replaced gi:13677055.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.hgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.2.  
Location/Qualifiers  
1..117583  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2313F11"

ORIGIN  
Query Match 37.0%; Score 740.6; DB 8; Length 117583;  
Best Local Similarity 83.7%; Pred. No. 4,4e-186;  
Matches 917; Conservative 0; Mismatches 129; Indels 49; Gaps 5;  
QY 1 AAAGCTTCTGAATGAGCAATGCTTCAACTCTTATACCAACTCTGAGTTGGG 60  
Db 104788 AAAGCTTCTGAATGAGCAATGCTTCAACTCTTATACCAACTCTGAGTTGGG 104729  
QY 61 GACATGGCTTCTCCCTTTCTAGGCTCTGTAAGCAAGTCTTCTGATTTGG 120  
Db 104728 GACATGGCTTCTCCCTTTCTAGGCTCTGTAAGCAAGTCTTCTGATTTGG 104669  
QY 121 GCCCTGATTTTAACTCTGCTGTAATTTCTTCTAGATGAGGCTCAAGT 180  
Db 104668 GCCCTGATTTTAACTCTGCTGTAATTTCTTCTAGATGAGGCTCAAGT 104609  
QY 181 ACAGATGATCTTCAATGATGACCCCAATGAGCTCAATCAACTCTGAGGAGC 240  
Db 104608 ACAGATGATCTTCAATGATGACCCCAATGAGCTCAATCAACTCTGAGGAGC 104553  
QY 241 CTGGAAGCAAGCCGCTGCTTCAATGAGCTCAATGAGCTCAATGAGCACTAC 300  
Db 104552 CTGGAAGCAAGCCGCTGCTTCAATGAGCTCAATGAGCTCAATGAGCACTAC 104493  
QY 301 CACTGAGAGGCTCTTCTTCACTCTATCAAGAGAGTCAAGTCAAGGCTATC-GCC 359  
Db 104492 CACTGAGAGGCTCTTCTTCACTCTATCAAGAGAGTCAAGTCAAGGCTATC 104433  
QY 360 AAATCCCAAGAGGCTGCTGCTTCTGAGGAGGAGTGAAGGAGGAGGAGGAGG 419  
Db 104432 AAATCCCAAGAGGCTGCTGCTTCTGAGGAGGAGTGAAGGAGGAGGAGGAGG 104373  
QY 420 GGGCTT-CTGGGTCAAGTGGGAGCTTGAAGAACTTTTGTCTAGCTAAAGATTTGAA 478  
Db 104372 GGGCTT-CTGGGTCAAGTGGGAGCTTGAAGAACTTTTGTCTAGCTAAAGATTTGAA 104313  
QY 479 TGCACCAATCAG-----CACTCT 496  
Db 104312 TGCACCAATCAGGCTTCAAGATTTGTAATGACCAATCCGCACTCTGTAATAAT 104253  
QY 497 GGTCTAGCTAAAGATTTGTAATGACCAATGAGCACTGTAATAATGAGCAATCAG 556  
Db 104252 GGTCTAGCTAAAGATTTGTAATGACCAATGAGCACTGTAATAATGAGCAATCAG 104193  
QY 557 AGGATGAGGCGGGGTCAATAGAGGATTAAGGCTGAGCAAGTGGGAGGAGGAGG 616  
Db 104192 AGGATGAGGCGGGGTCAATAGAGGATTAAGGCTGAGCAAGTGGGAGGAGGAGG 104133  
QY 617 CCCACTGGGCTCCCTTCCCACTGAGGAGCTTTGCTTTGCTTCAATTAATC 676  
Db 104132 CCCACTGGGCTCCCTTCCCACTGAGGAGCTTTGCTTTGCTTCAATTAATC 104073  
QY 677 TTGCTGCTGCTCAATTTTGTGTGCACTACTCTTATAGAGTGAAGTCACTGCGAG 736  
Db 104072 TTGCTGCTGCTCAATTTTGTGTGCACTACTCTTATAGAGTGAAGTCACTGCGAG 104013  
QY 737 GGTCTGAGCTTCAATCTGAGTCAAG-AGACCAAGAACCTGGAAGAAACAAGAA 795  
Db 104012 GGTCTGAGCTTCAATCTGAGTCAAG-AGACCAAGAACCTGGAAGAAACAAGAA 103953  
QY 796 GTCCGATGTGCTGCTTTAAGAGTGAAGTCACTGAGGAGCTTGTGAGCTTCACTC 855  
Db 103952 GTCCGATGTGCTGCTTTAAGAGTGAAGTCACTGAGGAGCTTGTGAGCTTCACTC 103893  
QY 856 CTGAAGTCAAGTGAAGCAACAAGCAAGAAAGAAAGTCTGAGCACTGGAATAT 915  
Db 103892 CTGAAGTCAAGTGAAGCAACAAGCAAGAAAGAAAGTCTGAGCACTGGAATAT 103833  
QY 916 CTGAAGAAACAAGTCAAGCAAGCAAGTCTTCAAGTCTGTAAGTCAAGTCAAGGAGGCT 975  
Db 103832 CTGAAGAAACAAGTCAAGCAAGCAAGTCTTCAAGTCTGTAAGTCAAGTCAAGGAGGCT 103773



Oy	976	1GTGGCTTCATCTTCTTGAATGTCAGCAAGACCAACCCACCGAAGAAACAATTCGAGA	1035
Db	103772	TGCCGCTTCATCTTCTTGAATGTCAGCGACCGAACCTACCAAGAAGCAATTCGGA	103773
Oy	1036	CACAGTAGAATCT	1050
Db	103712	CACAATTTCACATGCT	103698
RESULT 42			
AC091845/c			
LOCUS			
DEFINITION	AC091845	133322 bp	DNA linear HTG 09-JUN-2001
ACCESSION	AC091845		
VERSION	AC091845.1	GI:14333781	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
TITLE	Homidae; Homo.		
JOURNAL	1 (bases 1 to 133322)		
REFERENCE	DOE Joint Genome Institute.		
AUTHORS	Unpublished		
TITLE	Sequencing of Human Chromosome 5		
JOURNAL	2 (bases 1 to 133322)		
REFERENCE	DOE Joint Genome Institute.		
AUTHORS	Direct Submission		
TITLE	Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint		
JOURNAL	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	-----genome Center		
	Center: Joint Genome Institute		
	Center Code: JGI		
	Web site: <a href="http://www.jgi.doe.gov">http://www.jgi.doe.gov</a>		
	-----		
	Project Information		
	Center Project Name: 708368		
	Center clone name: CTRB-H1_2210H11		
	-----		
	Summary Statistics		
	Consensus quality: 129273 bases at least Q40		
	Consensus quality: 131004 bases at least Q30		
	Consensus quality: 131261 bases at least Q20		
	Estimated insert size: 134580; agarose-fp estimation		
	Estimated insert size: 132922; sum-of-contigs estimation		
	Quality coverage: 7.34 in Q20 bases; agarose-fp estimation		
	Quality coverage: 7.43 in Q20 bases; sum-of-contigs estimation.		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 5 'contigs'. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of 'N', but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	1. 133322		
	1253: contig of 1253 bp in length		
	1254		
	1353: gap of unknown length		
	1354		
	5163: contig of 3810 bp in length		
	5164		
	5263: gap of unknown length		
	5264		
	8261: contig of 2998 bp in length		
	8262		
	8361: gap of unknown length		
	8362		
	35118: contig of 26757 bp in length		
	35119		
	35218: gap of unknown length		
	35219		
	133322: contig of 98104 bp in length.		
	Location/Qualifiers		
	1. 133322		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="5"		
	/clone="CTD-2210H11"		

Query Match	37.0%	Score 740.6	DB 14	Length 133322
Best Local Similarity	83.7%	Pred. No. 4,4e-186		
Matches 917	Conservative	0	Mismatches 129	Indels 49
				Gaps 5
ORIGIN				
gap	/clone_lib="CalTech human BAC library D"			
gap	1254. .1353			
gap	/estimated_length=unknown			
gap	5164. .5263			
gap	/estimated_length=unknown			
gap	8262. .8361			
gap	/estimated_length=unknown			
gap	35119. .35218			
gap	/estimated_length=unknown			
1	AAAGGCTTCGAATCAGACAAATGCCCTTCAAACTCTTATACCAACTCTGAGTTGGGC	60		
26895	AAAGGCTTCGAATCAGACAAATGCCCTTCAAACTCTTATACCAACTCTGAGTTGGGT	26836		
61	GACATGACCTTCGCCCTTCTAGGCTCTGTGACAGCCATCTTGTATATGTCATTTGG	120		
26895	GACATGACCTTCGCCCTTCTAGGCTCTGTGACAGCCATCTTGTATATGTCATTTGG	26776		
121	GCCCTGATTTTAACTCTTGTGCAATTTGTTCTCTAGGATGAGAGCCATCAAGCT	180		
26775	GCCCTGATTTTAACTCTTGTGCAATTTGTTCTCTAGGATGAGAGCCATCAAGCT	26716		
181	ACAGATGATCTTCAAAATGTAACCCCAAAATGAGCTCACTTAACAATCTTGTGAGACC	240		
26715	ACAGATGATCTTCAAAATGTAACCCCAAAATGAGCTCACTTAACAATCTTGTGAGACC	26660		
241	CTGAGACCGACCGGCTGCGCTTTCATGCTCTAAAGAGCTCCCTCTGAGAGACATAC	300		
26659	CTGAGACCGACCGGCTGCGCTTTCATGCTCTAAAGAGCTCCCTCTGAGAGACATAC	26600		
301	CACGACGAGGCGCCCTTCTACCCCTATCCAGAGAGAATGAGCTACAGGCTCATC-GCC	359		
26599	CACTGACGAGGCGCCCTTCTACCCCTATCCAGAGAGAATGAGCTACAGGCTCATC-GCC	26540		
360	AAATCCCAACAGCAGCTGAGGCTGCTGTTTGGAGGGGAGATTGAGAGGTGAAGCCAGCT	419		
26539	AAATCCCAACAGCAGCTGAGGCTGCTGTTTGGAGGGGAGATTGAGAGGTGAAGCCAGCT	26480		
420	GAGGCTT-CTGGGTCAGGTGAGGAGACTTTTGTGTCTAGCTAAAGATTGTAA	478		
26479	GAGGCTT-CTGGGTCAGGTGAGGAGACTTTTGTGTCTAGCTAAAGATTGTAA	26420		
479	TGCACCATTCAG-----CACTCT	496		
26419	TGCACCATTCAGTCTCTAGCTAAAGATTGTAAATGACCAATCCGACTCTGTAAAT	26360		
497	GATGTACCTAAAGATTGTAAATGACCAATCAGCACTCTGTAAATGAGCAATCAGC	556		
26359	GATGTACCTAAAGATTGTAAATGACCAATCAGCACTCTGTAAATGAGCAATCAGC	26300		
557	AGATGTGAGGCGGAGGCTCAATTAAGGAGTAAATCTGAGCAAGCCAGCAGCTGAGCA	616		
26239	AGATGTGAGGCGGAGGCTCAATTAAGGAGTAAATCTGAGCAAGCCAGCAGCTGAGCA	26240		
617	CCCATTCGGGTCGCCCTTCCACATGCTGGAAGCTTTGTTCTTGTGCTTACAGATTAATC	676		
26239	CCCATTCGGGTCGCCCTTCCACATGCTGGAAGCTTTGTTCTTGTGCTTACAGATTAATC	26180		
677	TTGCTGAGCTCATCTTTGTGTGCTCACTACTTTATGAGCTGTAACTCTGAGG	736		
26179	TTGCTGAGCTCATCTTTGTGTGCTCACTACTTTATGAGCTGTAACTCTGAGG	26120		
737	GATGTGAGCTCATCTTTGTGTGCTCACTACTTTATGAGCTGTAACTCTGAGG	795		
26119	GATGTGAGCTCATCTTTGTGTGCTCACTACTTTATGAGCTGTAACTCTGAGG	26060		
796	CTCCGATGTGCTGCTTTAAGACTGTAACTCACTGGAAGCTCTGAGCTTCACTC	855		

Db 26059 CTCGGTCATGCCACTTTAAGACTGTAACTCACTGCGAAGCTCTGGGCTTCATC 26000  
 QY 856 CTGAAGTCAGTGTAGACCAACCAACCCACAGAGGAGAACTCTGGACACACCTGATAT 915  
 Db 25999 CTGAAGTCAGTGTAGACCCCAACCCCTGAGAGGAGAACTCTGGAGACATCTAAACAT 25940  
 QY 916 CTGAAGGAAACAACTCCAGACACACATCTTTGAGAGCTGTAACTACCTCAGCAAGGCTC 975  
 Db 25939 CTGAAGGAAACAACTCCAGACACACATCTTTAAGAACTGTAACTACCTGAGAGGCTC 25880  
 QY 976 TGTGGCTTCACTTTGTAAGTCAGCAAGACCAAGACCACCGAGAGGAAACAATTCAGA 1035  
 Db 25879 TGTGGCTTCACTTTGTAAGTCAGCAAGACCAAGACCACCAATTCAGCA 25820  
 QY 1036 CACAGTAGGAATCT 1050  
 Db 25819 CACAAATTAAACATGT 25805

RESULT 43  
 AL645568  
 LOCUS DEFINITION  
 AL645568 136031 bp DNA linear PRI 18-MAY-2005  
 Human DNA sequence from clone RP11-296014 on chromosome 1. Contains a pseudogene similar to part of ribosomal protein L26 (RP126), a novel pseudogene and a novel gene, complete sequence.

ACCESSION  
 AL645568 GI:18072559  
 VERSION  
 HTG; RP126.  
 KEYWORDS  
 Homo sapiens (human)  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 136031)  
 Matthews,N.  
 REFERENCE  
 Direct Submission  
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
 Clonerequest: clonerequest@sanger.ac.uk  
 On Jan 6, 2002 this sequence version replaced gi:17977921.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chri  
 RP11-296014 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: vegas@sanger.ac.uk  
 -----

FEATURES  
 source  
 1..136031  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-296014"

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 Location/Qualifiers

misc\_feature  
 1  
 /clone\_11b="RP11-11.2"  
 /note="Clone Left\_end: RP11-296014"  
 11587..11902  
 /locus\_tag="RP11-296014.1-001"  
 /pseudo  
 11587..11902  
 /locus\_tag="RP11-296014.1-001"  
 /note="match: proteins: Tr:BAC97863 Tr:Q14704 Tr:Q810F2 Sw:Q8U9N3 Tr:AAH17360 Tr:BAC21653 Tr:BAC56565 Tr:O7SXA1 Tr:Q862N4 Tr:Q90YU2"  
 /pseudo  
 /product="pseudogene similar to part of ribosomal protein L26 (RP126)"  
 complement(22296..23631)  
 /locus\_tag="RP11-296014.2-001"  
 /pseudo  
 complement(22296..23631)  
 /locus\_tag="RP11-296014.2-001"  
 CDS  
 complement(22296..23631)  
 /locus\_tag="RP11-296014.2-001"  
 /note="match: proteins: Tr:BAC97863 Tr:Q14704 Tr:Q810F2 Tr:Q865V0 Tr:Q86XR8 Tr:Q8CEB0 Tr:Q81XP0 Tr:Q9BVF9 Tr:Q9D5W5"  
 /pseudo  
 /codon\_start=1  
 /product="novel pseudogene"  
 34722..34863  
 /note="Sequence from overlapping clone RP11-192B4 (AL360002). Assembly confirmed by restriction digest."  
 60074  
 /note="Clone right end: RP11-192B4"  
 complement(join(77691..78520,120276..121260))  
 /locus\_tag="RP11-296014.3-001"  
 complement(join(77691..78520,120276..121260))  
 /locus\_tag="RP11-296014.3-001"  
 /product="novel transcript"  
 /note="match: ESTs: Em:AA743066.1 Em:AA806822.1 Em:AA812123.1 Em:AF655890.1 Em:BX117627.1 match: cDNAs: Em:AK130460.1"  
 complement(77691)  
 /locus\_tag="RP11-296014.3-001"  
 complement(77713..77718)  
 /locus\_tag="RP11-296014.3-001"  
 complement(join(77972..78520,100361..100648,101817..101995))  
 /locus\_tag="RP11-296014.3-002"  
 complement(join(77972..78520,100361..100648,101817..101995))  
 /locus\_tag="RP11-296014.3-002"  
 /product="novel transcript"  
 /note="match: ESTs: Em:AU521783.2 Em:BQ222905.1 Em:BX345040.1"  
 134032  
 /note="Clone Left\_end: RP3-436N22"

ORIGIN  
 Query Match 37.0%; Score 740.4; DB 8; Length 136031;  
 Best Local Similarity 82.0%; Pred. No. 5e-186;  
 Matches 945; Conservative 0; Mismatches 131; Indels 76; Gaps 5;

1 AAAGGCTTGAAATAGACANTGCTTCAACCTTTATACCAACCTTGAGTGGG 60  
 Db 41251 AAAGGCTTGAAATAGACANTGCTTCAACCTTTATACCAACCTTGAGTGGG 41310  
 QY 61 GACATGGCTTCCCTCTTCTAGAGTCCTGTGACAGCATCTTGCTATAGTGCATTGG 120  
 Db 41311 AACATGGCTTCCCTCTTCTAGAGTCCTGTGACAGCATCTTGCTATAGTGCATTGG 41370  
 QY 121 GCCCTGATTTTAACTTGTGCAATTTGTTCTTGTAGATGAGCCATCAAGCT 180  
 Db 41371 GCCCTGATTTTAACTTGTGCAATTTGTTCTTGTAGATGAGCCATCAAGCT 41430  
 QY 181 ACAGATGATCTTAAATGTAACCCCAATGAGCTCAATCAACTTGTGAGAGACC 240

```

Db      41431 ACAGACGGCTTACAAATGGAACCCCAATGAGCTCACTAACAACCTTCTACCAAGAAC 41490
OY      241 CCTGACCGACCCCGCTGCGCCCTTCAATGAGCCCTAAAGAGCTCCCTCGGAGAGACACTAC 300
Db      41491 CTTGAACCAACCACTATGGCTTTTCACTGGCTTAAAGATTCCTCTGGAGACACTAC 41550
OY      301 CACTCAGAGGCGCCCTTCTTCAACCCCTATCCAGAGGAATAGCTACAGCGCTATCG-CC 359
Db      41551 AACTCAGAGGCTTCTTCTTCAACCCCTATCCAGAGGAATAGCTACAGCGCTATCGCC 41610
OY      360 AATCCCAACAGACAGCTGGGCTGTCTGTTTGGAGGGGAGATTGAGAGTGAAGCCAGCT 419
Db      41611 AATTCACCAACAGACAGCTGGGCTGTCTGTTTGGAGGGGAGATTGAGAGTGAAGCCAGCT 41670
OY      420 GGGCTTTCTGGGCTCAGGTGGGAGCTTGGAGAACTTTGTCTTACAGCTTAAAGATTGT- 475
Db      41671 GGGCTTTCTGGGCTCAGGTGGGAGCTTGGAGAACTTTGTCTTACAGCTTAAAGATTGTAAAC 41730
OY      476 ----- 475
Db      41731 AACCAATCAGGCTCTGTGTCTAGCTAAAGTTTGTAAATGCAACCAAGACACTGTGTA 41790
OY      476 -AAATGCAACCAATCAGCACTGTGTGTCTAGCTAAAGATTGTAAATGCAACCAATGACAC 534
Db      41791 AAATGCAACCAATCAGCTGTGTGTCTAGCTAAAGATTGTAAATGCAACCAATGACAGTC 41850
OY      535 TCTGTAAATGCAACCAATCAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 594
Db      41851 TCTGTAAATGCAACCAATCAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 41910
OY      595 CCACCCGAGCAGCAGTGGGCAACCCACTCGGGTCCCTTCCACACTGTGTGTGTGTGTGTGTGT 654
Db      41911 CCACTGAGCCAGCAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 41970
OY      655 CTTTGTCTCTTCAAAATTAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 714
Db      41971 CTTTGTCTCTTCAAAATTAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 42030
OY      715 GAGCTTAACTCACTGCGAGGGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 773
Db      42031 GAGCTTAACTCACTGCGAGGGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 42090
OY      774 ACCCAGTGAAGGAACAAGAACTCCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 833
Db      42091 ACCCAGTGAAGGAACAAGAACTCCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 42150
OY      834 GCGAAGCTCTGAGCTTCACTCTGTGAATC-AGTGAAGCAACAACCCACAGAGGAG 892
Db      42151 GCAAAGTCTGTGGCTTCACTCTGTGAATCAAGCTAGACACAGAACCCACAGAAATGAG 42210
OY      893 AAATCTGTGACACACTGAATATCTGAAGGAACAACCTCCAGACACACTTTTTCAG 952
Db      42211 AAATCTGTGACAC-----ATCTGAAGGAACAAGCTCCGACACACACTTTTTCAG 42262
OY      953 CTGTAAACACTCAGCGAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1012
Db      42263 CTGTAAACACTCAGCGAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 42322
OY      1013 CACCGAAGGAACAATTCAGACACAGTGAAGGAATCTGTATTTTGTGTGTGTGTGTGTGTGT 1072
Db      42323 CACCGAAGGAACAATTCAGACACAGTGAAGGAATCTGTATGTATGTATGAGCGGTTG 42382
OY      1073 AGGGTTACTCCA 1084
Db      42383 ATGAGTGCAGCA 42394

```

RESULT 44  
 AL591074 203076 bp DNA linear PRI 18-MAY-2005  
 LOCUS Human DNA sequence from clone XXYac-60D10 on chromosome 20 Contains  
 DEFINITION the CSTs gene for cystatin D, the CST2 gene for cystatin SA, the

```

ACCESSION      AL591074
VERSION        1.0
KEYWORDS       HTG; CST1; CST2; cystatin D; cystatin SA; cystatin SN.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Homidae; Homo.
REFERENCE      1 (bases 1 to 203076)
AUTHORS       Phillimore, B.
TITLE         Direct Submision
JOURNAL       Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgehire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT       On Oct 2, 2001 this sequence version replaced gi:15705266.
                The following abbreviations are used to associate primary accession
                numbers given in the feature table with their source databases:
                Em; EMBL; Sw; SWISSPROT; Tr; TrEMBL; Wp; WormPeP; Information
                on the WormPeP database can be found at
                http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
                was generated from part of bacterial clone contigs of human
                chromosome 20, constructed by the Sanger Centre Chromosome 20
                Mapping Group. Further information can be found at
                http://www.sanger.ac.uk/HGP/Chr20
                XXYac-60D10 is from a ICRP YAC library
                VECTOR: pYAC4
                ----- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: vegas@sanger.ac.uk
                -----
                This sequence was finished as follows unless otherwise noted: all
                regions were either double-stranded or sequenced with an alternate
                chemistry or covered by high quality data (i.e., phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats; all regions were covered by at least
                one subclone; and the assembly was confirmed by restriction digest,
                except on the rare occasion of the clone being a YAC.
                Location/Qualifiers
                source          1..203076
                                /organism="Homo sapiens"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:9606"
                                /chromosome="20"
                                /clone="XXYac-60D10"
                                /clone_11b="Human-ICRF-YAC"
                                /note="Clone right end: RP1-234M6"
                                complement(1587..1673)
                                /locus_tag="AL591074.2-002"
                                /pseudo
                                complement(1587..1673)
                                /locus_tag="AL591074.2-002"
                                /note="match: proteins: AAH21225 O19092 O19093 P01036
                                P01037 P09228"
                                /pseudo
                                /codon_start=1
                                1919..2301
                                /note="Single clone region. Sequence from reads from a
                                short insert library derived form a clone PCR."
                                14801..15094
                                /note="Single clone region. Sequence from reads from a
                                short library derived from a clone PCR."
                                15422..15653
                                /note="Single clone region. Sequence from reads from a
                                short insert library derived from a clone PCR."
                                complement(37829..41204)
                                /gene="CST1"
                                /locus_tag="AL591074.1-001"
                                complement(join(37829..38166,39283..39396,40906..41204))

```

/gene="CST1"  
/locus\_tag="AL591074.1-001"  
/note="match: ESTs: BF689725 BF689815 BG287350 BG290010  
BG290406 BG291293 BG403032 BG403934 BG765263  
match: CDNAS: BC021225.1 J03870.1 X54667.1"  
complement(37829)  
/gene="CST1"  
/locus\_tag="AL591074.1-001"  
complement(join(38083..38166,39283..39396,40906..41133))  
/gene="CST1"  
/locus\_tag="AL591074.1-001"  
/standard\_name="OTTHUMP0000030444"  
/note="match: proteins: CAC07196 P01035 P01036 P01037"  
/codon\_start=1  
/protein\_id="CAC94783.1"  
/db\_xref="GI:16116524"  
/db\_xref="GeneW:2473"  
/db\_xref="GOA:P01037"  
/db\_xref="HSSP:IG96"  
/translation="MAOYLSTLLILATLAVLAWSPKPERDRIIPGIYNADIDEMV  
QALHRAISRYNKATKDDYRRRLRLRAQCIVGVNYPFDIEVGRITCTKSQPMID  
TCAHFQPELQKQKQCSFELYEVPMERNRSLVSRQCS"  
73978..74211  
/note="Single clone region. Sequence from reads from a  
short insert library derived from a clone PCR."  
90914..91427  
/note="Single clone region. Sequence from reads from a  
short insert library derived from a clone PCR."  
complement(114039..116942)  
/gene="CST2"  
/locus\_tag="AL591074.3-001"  
complement(join(114039..114370,115477..115590,  
116700..116942))  
/gene="CST2"  
/locus\_tag="AL591074.3-001"  
/note="match: ESTs: AA523539 A1521546 BG700542 BG706328  
BG708038 B1488523 B1549132 B1551519 B1601985  
match: CDNAS: AF311741.1 BC013083.1 U52028 X05607  
X05607.1"  
complement(114039)  
/gene="CST2"  
/locus\_tag="AL591074.3-001"  
complement(114054)  
/gene="CST2"  
/locus\_tag="AL591074.3-001"  
complement(join(114287..114370,115477..115590,  
116700..116927))  
/gene="CST2"  
/locus\_tag="AL591074.3-001"  
/standard\_name="OTTHUMP0000030445"  
/note="match: proteins: AAK1570 P09228"  
/codon\_start=1  
/protein\_id="CAC94784.1"  
/db\_xref="GI:16116525"  
/db\_xref="GeneW:2474"  
/db\_xref="GOA:P09228"  
/db\_xref="HSSP:IG96"  
/translation="MAMPLCTLLILATLAVLAWSPKPERDRIIPGIYNADIDEMV  
QALHRAISRYNKATEDYRRRLRLRAREQIVGVNYPFDIEVGRITCTKSQPMID  
TCAHFQPELQKQKQCSFELYEVPMERNRSLVNSRCOA"  
complement(166423..169947)  
/gene="CST5"  
/locus\_tag="AL591074.5-001"  
complement(join(166423..166538,167772..167885,  
169713..169947))  
/gene="CST5"  
/locus\_tag="AL591074.5-001"  
/note="match: CDNAS: X70377"  
complement(join(166455..166538,167772..167885,  
169713..169943))

/gene="CST5"  
/locus\_tag="AL591074.5-001"  
/standard\_name="OTTHUMP0000030448"  
/note="match: proteins: P28325"  
/codon\_start=1  
/protein\_id="CA123644.1"  
/db\_xref="GI:56205430"  
/db\_xref="InterPro:IPR000010"  
/translation="MMMPHETPLRLITAMVAVGASASQSRLLAGIHAITDNDKSV  
QCALDPAISRYNKVINDYRSPLPLQVMAVQIVGVNYPFVVKGRITCTKSQPMID  
DNCPEMDQPLKEERCSFQINSEPMEDKISILNYCRKY"  
181333..181442  
/note="Single clone region. Sequence from reads from a  
short insert library derived from a clone PCR."  
202977  
/note="Clone\_left\_end: RP11-1P13"  
ORIGIN  
Query Match 37.0%; Score 739.8; DB 8; Length 203076;  
Best Local Similarity 85.7%; Pred. No. 7.2e-186;  
Matches 914; Conservative 0; Mismatches 122; Indels 31; Gaps 7;  
misc\_feature  
misc\_feature  
misc\_feature  
1 AAAGGCTTGAATAGACAAATGCTTCAAACTTTATACCACTCTGAGTTGGC 60  
31970 AAACGATTCGAATCAGATTAACACCTTCAAGCTTTATACCACTCTGAGTTGAGC 32029  
61 GACATGGCTTCTCCCTTTAGAGTCTGTGACAGCCATTTGCTAATAGTCATTGG 120  
33030 AACATGGCTTCTCCCTTTAGAGTACCCGAGCAGCATCTTCTAATACCTCCCTCGG 33089  
121 GCCCTGATTTTAACTCTGTGTCAAATTTGTTCTCTAGAGATGAGGCCATCAAGCT 180  
32090 GCCCTGATTTTAACTCTGTGTCAAATTTGTTCTCTAGAGATGAGGCCATCAAGCT 32149  
181 ACAGATGATCTTCAATATGAATGAACCCAAATGAGCTCACTTAACATCTGTGAGAGCC 240  
32150 TCGATGATCTTCAATATGAATGAACCCAAATGAGCTCACTTAACATCTGTGAGAGACC 32209  
241 CTGAGACGACCGCGTGGCCCTTTCAATGAGCTCCCTCTGAGAGACATAC 300  
32210 CTGAGACATCAACCACTGCGCCCTTTCACTGAGCTTAAAGATTCACCTGTGAGATAC 32269  
301 CACTGAGAGGCCCTCTTCAACCCCTATCAGAGAGAAAGTACAGCGCTATC-GCC 359  
32270 AACTGAGAGGCCCTCTTCTTCCCTATCAGAGAGAAAGTACAGCGCTATCAGTTG 32329  
360 AATCCCAACAGAGCTGGGGTGTCTGTTTGGAGGGGGAATTGAGAGTGAAGCCAGCT 419  
32330 AACTCCCAACAGAGTGGGGGTGTCTGTTT-GAGAGGGATTTGACAGGTGAAGCCAGCT 32388  
241 CTTGAGACGACCGCGTGGCCCTTTCAATGAGCTCCCTCTGAGAGACATAC 300  
32210 CTGAGACATCAACCACTGCGCCCTTTCACTGAGCTTAAAGATTCACCTGTGAGATAC 32269  
301 CACTGAGAGGCCCTCTTCAACCCCTATCAGAGAGAAAGTACAGCGCTATC-GCC 359  
32270 AACTGAGAGGCCCTCTTCTTCCCTATCAGAGAGAAAGTACAGCGCTATCAGTTG 32329  
360 AATCCCAACAGAGCTGGGGTGTCTGTTTGGAGGGGGAATTGAGAGTGAAGCCAGCT 419  
32330 AACTCCCAACAGAGTGGGGGTGTCTGTTT-GAGAGGGATTTGACAGGTGAAGCCAGCT 32388  
420 GGGCTTCTGGGTGAGGTGGGGAATTGAGAGCTTTGTGCTAAGTAAAGATTGTAAT 479  
32389 AGGCTTCTGGGTGAGGTGGGGAATTGAGAGCTTTGTGCTAAGTAAAGATTGTAAT 32448  
480 GCACCAATCAGACCTGTGTCTAGCTAATAAGATTGTAATGACC-----525  
32449 GCACCAATCAGACCTGTGTCTAGCTAATAAGATTGTAATGAGCACTGTGT 32508  
526 -----AATCAGACCTGTAAATGAGCAATCAGAGAGTGGGGGGGTCA 574  
32509 AAAACAGACTAATCAGACCTGTAAATGAGCAATCAGAGAGTGGGGGGGGCCCA 32568  
575 AATTAAGGAGTAAATACTGGCCACCCGAGCCAGAGTGGCAACCACTGGGTCCCCTTC 634  
32569 AATTAAGGAGTAAATACTGGCCCA-CCTAAGCAGAGTGGCAACCACTGGGTCCCCTTC 32627  
635 CACACGTGGAAGCTTGTCTTTTGTCTTGTCTTCAATTAATCTGTGCTGCTCACTTCT 694  
32628 CACGCTGTGGAAGCTTGTCTTTTGTCTTGTCTTCAATTAATCTGTGCTGCTCACTTCT 32687  
695 TGTGTCCACACTTACTTTATGAGCTGTAACTCACTGAGAGGCTGTGGCTTCACTTCC 754  
32688 TGGGTCTCACTTACTTTATGAGCTGTAACTCACTGAGAGGCTGTGAGCTTCACTCC 32747

OY	755	TGAAGTCA -ACAGCCACGAAACCACCTGGAAAGGAACAAGAATCCTCCGAGTCCTCTT	813
Db	32748	TGAAGTCAGTGAGGACACAAAACCCACGAGGGAATGACAACTCCAGACATCAC-C-T	32806
OY	814	TAAAGCTGTAAACTCACTGCAGAGCTGTGCAGACTTCACTCCTGTAAGTAGTGACCA	873
Db	32807	TAAAGCTGTAAACTCACTGTGTGAAGGCTGTGGCTTCACTTCTGTAAGTAGCAAGCA	32866
OY	874	CAAACCCACCGAAGAAAGAACTCTGACACACCTTGAAATATCTGAAAGAAACAATTCCA	933
Db	32867	CGAACCCACCGAAGAAAGAAATCTCCAGACATAATATCAATCTGAAAGAAC-AACTCTG	32922
OY	934	GACACACCATTTTTCAGAGCTGTAAACACTCACCCGCAAGGGCTGTGGCTTCATTCTTGA	993
Db	32926	GACACACCATTTTTCAGAGCTGTAAACACTCACCCGCAAGGGCTGTGGCTTCATTCTTGA	32985
OY	994	GTCCAGCAAGACCAAGAACCCACCGAAGAAACAATTCAGACACAG	1040
Db	32986	GTCCGCCAGGCGCAAGAACCCACCGAAGAAACAATTCAGACACAG	33032
RESULT_45	AC026307/C	LOCUS	
DEFINITION	Homo sapiens 12 BAC Rpl1-753N [Roswell Park Cancer Institute Human BAC library] complete sequence.		
ACCESSION	AC026307		
VERSION	AC026307.16		
KEYWORDS	GI:12621201		
SOURCE	HTG.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 16357) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarta,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowley,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C., Butch,P., Burkett,C., Burrill,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chin,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Garroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H., Doutheaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emertling,S., Escotlo,M., Falls,T., Ferriguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gablis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulik,S., Hunne,J., Ioshikius,I., Jackson,L.E., Jacobson,B., Jela,Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvah,J., Kovac,C., Kravovic,J., Kurehi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,C., Liu,W., Loulsegad,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Matindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Merschler,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabab,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenko,S., Ogun,M., Okunnu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primm,E., Pu,L.B., Quiles,M., Renz,Y., Rives,M., Rojås,S., Scott,G., Shen,H., Shm,C., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisom,I., Sodestegren,E., Sonikeit,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swatek,A., Tabor,P., Tametisa,A., Tametisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tafirof,B., Thomas.N., Thomas.S., Uemant,K., Vasquez,L., Vera,V., Villallon,D.		

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITILE
JOURNAL
REFERENCE
AUTHORS
TITILE
JOURNAL
COMMENT

Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watlington,S., Williams,G., Williamson,A.,
Wleczyk,K.R., Wooden,S., Morley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorrilla,S., Kucheriapatti,R. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 166357)
Morley,K.C.
Submitted (22-Mar-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 166357)
Morley,K.C.
Direct Submission
Submitted (31-Jan-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 166357)
Morley,K.C.
Direct Submission
Submitted (01-Feb-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
ON Jan 31, 2001 this sequence version replaced gi:12585011.
INFORMATION: http://www.bgsc.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarly (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT-----

Config Length: Summary Statistics -----
Phrap values in estimate: 166357
Average error rate (BCM-Phrap estimate): 15586
Fraction of Phrap values less than 40 : 7.20476e-06
Number of consensus changing edits: 0.00725507
Number of N's in consenus : 184
0
```



```

OY      905  CACCTGAATATCTGAAGGAAACAACCTCCAGACACACATCTTTAGAGCTGTAACTCA  964
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      14644  CATCTGAACATCTGAAAGAACAACTCCAGACACACCATCTTTAAGAACTGTAACTCA  146585
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      965  CCGCAAGGGTCTGTGGCTTCATTTCTTGAAGTCAGCAAGACCAAGAACCCACCGGAAGGAA  1024
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      146584  CCACAAGGGTTCGTGGCTTCATTTCTTGAAGTCAGTGAGACCAAGAACCCACCGGAAGGAA  146525
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      1025  CAAATTCAGACACAGTAGGAATCTGTATTTTGAT  1061
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      146524  CCAATTCGAGACACAGAACTAGTATCAATATTAAAT  146488
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: January 22, 2006, 13:38:37  
 Job time : 6846.68 secs





93	363.8	18.2	110000	4	AAK95240.00	AAK95240 Human neu
94	363.8	18.2	110000	4	AAK96733.00	AAK96733 Human neu
95	363.8	18.2	110000	6	ABT00010.00	ABT00010 Human neu
96	363.8	18.2	110000	6	ABT01503.00	ABT01503 Human neu
97	363.8	18.2	110000	11	ADM70291.00	ADM70291 Human neu
98	363.8	18.2	110000	12	ADH7486.00	ADH7486 Human neu
99	361.8	18.1	1043	3	AAZ7494	AAZ7494 Human 5'
100	360	18.0	231222	10	ADL13693	ADL13693 Osteoarthritis

ALIGNMENTS

RESULT 1

AAV83939 standard; DNA; 80595 BP.

AAV83939; 03-MAR-1999 (first entry)

HC-contig derived from normal human chromosome 10q25.2 region.

yeast artificial chromosome; YAC; probe; eukaryotic chromosome; neocentromere; replication; extra-chromosomal element; segregation; cell division; artificial chromosome; gene therapy; mardel(10); human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.

Homo sapiens.

MO9851790-A1.

19-NOV-1998.

13-MAY-1998; 98WO-AU000352.

13-MAY-1997; 97AU-00006784.

26-AUG-1997; 97AU-00008791.

(AMRA-) AMRAD OPERATIONS PTY LTD.

Choo K, Du Sart D, Cancilla MR;

WPI, 1999-009773/01.

New isolated nucleic acid comprising neocentromere sequences from eukaryotic chromosome - used to produce replicable, segregating artificial chromosomes that can carry large amounts of DNA for gene therapy.

Claim 8; Fig 6; 540pp; English.

The present sequence represents the HC-contig derived from normal human chromosome 10, 10q25.2 region. This region can be naturally mutated to produce an unusual chromosomal marker referred to as mardel(10). The mardel(10) marker is mitotically stable and contains a functional neocentromere at a location regarded as non-centromeric. This neocentromere maps to q25.2 on chromosome 10. The specification describes nucleic acid sequences derived from a eukaryotic chromosome, including a neocentromere or its functional derivative or hybrid, that are able, in a compatible cell, of replicating, acting as extra-chromosomal element and segregating during cell division. The sequences can be used to construct artificial chromosomes for use in gene therapy comprising a replicable, segregating nucleic acid that confers a specific phenotype on cells. Human artificial chromosomes can propagate in human cells and carry large amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal, they are not mutagenic. The artificial chromosomes are also useful for generation of transgenic plants and animals, in production of cytokines, and to make diagnostic reagents, e.g. for expression of cytokines, receptors and growth factors, or to increase the copy number of a gene in a cell. The constructs may also be used for functional and structural analysis of chromosomes

SQ	Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;
Query Match	100.0%; Score 2001; DB 2; Length 80595;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2001; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 AAAAGCTTGAATACGACAAATGCTTCAAACTCTTATACCAACCTCTGAGTTGGCC 60
DB	59000 AAAAGCTTGAATACGACAAATGCTTCAAACTCTTATACCAACCTCTGAGTTGGCC 59059
QY	61 GACATGGCTTCCCTCTTCTAGGTCCTGTGACAGCAATCTTGTAATAGTGCATTTGG 120
DB	59060 GACATGGCTTCCCTCTTCTAGGTCCTGTGACAGCAATCTTGTAATAGTGCATTTGG 59119
QY	121 GCCCTGATTTTAACTCTTGTCGTAATTTGTTCTCTAGATCGAGGCATCAAGCT 180
DB	59120 GCCCTGATTTTAACTCTTGTCGTAATTTGTTCTCTAGATCGAGGCATCAAGCT 59179
QY	181 ACAGATGATCTTAACAAATGTAAACCCCAATGAGCTCACTAACAACTTGTGAGGAGCC 240
DB	59180 ACAGATGATCTTAACAAATGTAAACCCCAATGAGCTCACTAACAACTTGTGAGGAGCC 59239
QY	241 CCTGACGACCGGCTGGCCCTTTCATGAGCCCTAAAGAGCTCCCTCTGAGGACACTAC 300
DB	59240 CCTGACGACCGGCTGGCCCTTTCATGAGCCCTAAAGAGCTCCCTCTGAGGACACTAC 59299
QY	301 CACTGAGGAGCCCTTCTTCAACCCCTATCAGACAGGAATGACTACAGCGGTATCGCA 360
DB	59300 CACTGAGGAGCCCTTCTTCAACCCCTATCAGACAGGAATGACTACAGCGGTATCGCA 59359
QY	361 AATCCCAACAGACGCTGGGGTGTCTGTTTGGAGGGGGATTTGAGAGGTGAAGCCGACG 420
DB	59360 AATCCCAACAGACGCTGGGGTGTCTGTTTGGAGGGGGATTTGAGAGGTGAAGCCGACG 59419
QY	421 GGCTTGGGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
DB	59420 GGCTTGGGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 59479
QY	481 CACCAATCAGACCTGTGTCTAGCTAAAGATTTGTAATGACCAATCAGACCTGTGTGA 540
DB	59480 CACCAATCAGACCTGTGTCTAGCTAAAGATTTGTAATGACCAATCAGACCTGTGTGA 59539
QY	541 AAATGACCAATCAGACGAGATGTGGCGGGGCTCAATTAAGGAGTAAAACTGGCCACCC 600
DB	59540 AAATGACCAATCAGACGAGATGTGGCGGGGCTCAATTAAGGAGTAAAACTGGCCACCC 59599
QY	601 GAGCCAGACGTGGCAACCCACTGGGATCCCTCCACACTGTGGAAGCTTGTCTTTTG 660
DB	59600 GAGCCAGACGTGGCAACCCACTGGGATCCCTCCACACTGTGGAAGCTTGTCTTTTG 59659
QY	661 CTCTTCAATTAATCTTGTGCTGCTCAATCTTGTGTCCACACTAATCTTATGAGCTG 720
DB	59660 CTCTTCAATTAATCTTGTGCTGCTCAATCTTGTGTCCACACTAATCTTATGAGCTG 59719
QY	721 TAACTACTACTGCGAGGGTGTGGCTTCAATCTTGAAGTCAACAGACCAACCTCACT 780
DB	59720 TAACTACTACTGCGAGGGTGTGGCTTCAATCTTGAAGTCAACAGACCAACCTCACT 59779
QY	781 GGAAGAAACAAAGAACTCCGATGTGCTTTAAGAGTGAACCTCACTGCGAAGC 840
DB	59780 GGAAGAAACAAAGAACTCCGATGTGCTTTAAGAGTGAACCTCACTGCGAAGC 59839
QY	841 TTGACAGCTTCACTCTGAGAGTCAAGTGAACCAACCAACCAACCAACCAACCAACCAAC 900
DB	59840 TTGACAGCTTCACTCTGAGAGTCAAGTGAACCAACCAACCAACCAACCAACCAACCAAC 59899
QY	901 GACACACCTGAATATCTGAAGAAACAACTTCACAGACACACATCTTCAAGCTGTAA 960
DB	59900 GACACACCTGAATATCTGAAGAAACAACTTCACAGACACACATCTTCAAGCTGTAA 59959
QY	961 CTACCGCAAGGCTGTGCTTCAATCTTGAAGTCAAGCAAGCAAGCAAGCAAGCAAGCA 1020
DB	59960 CTACCGCAAGGCTGTGCTTCAATCTTGAAGTCAAGCAAGCAAGCAAGCAAGCAAGCA 60019

```
OY 1021 GGAACAATTCAGACAGACAGTAAATCTGATTTTGTGCTGGCTCCAGGGTTAC 1080
    |||
    |||
    |||
Db 60020 GGAACAATTCAGACAGACAGTAAATCTGATTTTGTGCTGGCTCCAGGGTTAC 60079
OY 1081 TCCAGTCATTTGAAGTCTCCATTCAGTCCCTTAAGAGAAACAGAAATGTTTGAAGAGCAC 1140
    |||
    |||
    |||
Db 60080 TCCAGTCATTTGAAGTCTCCATTCAGTCCCTTAAGAGAAACAGAAATGTTTGAAGAGCAC 60139
OY 1141 ATGTGGAAATTTGTTATGAGACAGAGCTTGAATGACATAGAGGATTTTGTATCAAACTTA 1200
    |||
    |||
    |||
Db 60140 ATGTGGAAATTTGTTATGAGACAGAGCTTGAATGACATAGAGGATTTTGTATCAAACTTA 60199
OY 1201 GCTGAGAGAGGGCCAGAGAAATTAATCTAAGAGAGCAGTTTGTAGACAGTAGATAGT 1260
    |||
    |||
    |||
Db 60200 GCTGAGAGAGGGCCAGAGAAATTAATCTAAGAGAGCAGTTTGTAGACAGTAGATAGT 60259
OY 1261 CTTTGATCTGAGACATGATGATTAATCAAGCAATTAATTAAGAAAAATATAGCCAGGTGC 1320
    |||
    |||
    |||
Db 60260 CTTTGATCTGAGACATGATGATTAATCAAGCAATTAATTAAGAAAAATATAGCCAGGTGC 60319
OY 1321 GATGGCTCATGCTGTGATATCCAGCACTTTGGAGGCCAAGGGGTGTGATCAGAGGTTC 1380
    |||
    |||
    |||
Db 60320 GATGGCTCATGCTGTGATATCCAGCACTTTGGAGGCCAAGGGGTGTGATCAGAGGTTC 60379
OY 1381 AGGCGTTGAGACACAGCTGGCCCAATGAGTGAACCCCGTCTTAATAAAATACAAAA 1440
    |||
    |||
    |||
Db 60380 AGGCGTTGAGACACAGCTGGCCCAATGAGTGAACCCCGTCTTAATAAAATACAAAA 60439
OY 1441 ATTAGCGTGTGTGTGAGCAGCATCTGTAATCCAGTACTCAGAGGCTGAGGACAGGG 1500
    |||
    |||
    |||
Db 60440 ATTAGCGTGTGTGTGAGCAGCATCTGTAATCCAGTACTCAGAGGCTGAGGACAGGG 60499
OY 1501 AATCTCTGAACCTTGGAGGACAGAGGTTGCACTGAGCCAAAGTACACACACAGCACTCA 1560
    |||
    |||
    |||
Db 60500 AATCTCTGAACCTTGGAGGACAGAGGTTGCACTGAGCCAAAGTACACACACAGCACTCA 60559
OY 1561 TCTCGGTGACAGAGGACATCTGCTCAAAAAAAGAAAAAAGAAAAAAGAAAAA 1620
    |||
    |||
    |||
Db 60560 TCTCGGTGACAGAGGACATCTGCTCAAAAAAAGAAAAAAGAAAAAAGAAAAA 60619
OY 1621 AATATCAAGAAATTTGACAGGTAACTTATTAACACTTATCTATGACACAGCAATACA 1680
    |||
    |||
    |||
Db 60620 AATATCAAGAAATTTGACAGGTAACTTATTAACACTTATCTATGACACAGCAATACA 60679
OY 1681 CTAAGTGTTTACATGATTAATCTAATCTTAACAATAGCCCTATGAGTCAAGTGC 1740
    |||
    |||
    |||
Db 60680 CTAAGTGTTTACATGATTAATCTAATCTTAACAATAGCCCTATGAGTCAAGTGC 60739
OY 1741 TGTATTAATCTCACTTATTAAGTAAGAACTGAAGTACAGAAAGTCAAGTAAGAA 1800
    |||
    |||
    |||
Db 60740 TGTATTAATCTCACTTATTAAGTAAGAACTGAAGTACAGAAAGTCAAGTAAGAA 60799
OY 1801 TGGCCATGCTTCATCTCAGTTTGAAGCAACTCTTACAGAACTCTGCTGAGAAAT 1860
    |||
    |||
    |||
Db 60800 TGGCCATGCTTCATCTCAGTTTGAAGCAACTCTTACAGAACTCTGCTGAGAAAT 60859
OY 1861 GCTCTAACAGATGTGAGTCAAGGGTGTGAGAGTACTGAGTGTGAGGAGTTGGG 1920
    |||
    |||
    |||
Db 60860 GCTCTAACAGATGTGAGTCAAGGGTGTGAGAGTACTGAGTGTGAGGAGTTGGG 60919
OY 1921 ATGGAAGATGTGATGAAGAAACAGCTTGAACAGAAAGTCAACTCTCTGTGGGA 1980
    |||
    |||
    |||
Db 60920 ATGGAAGATGTGATGAAGAAACAGCTTGAACAGAAAGTCAACTCTCTGTGGGA 60979
OY 1981 CCTTGAAGGTTAGAGGACT 2001
    |||
    |||
    |||
Db 60980 CCTTGAAGGTTAGAGGACT 61000
    |||
    |||
    |||
```

RESULT 2  
AAV83940  
ID AAV83940 standard; DNA; 80240 BP.

```
XX AAV83940;
AC 03-MAR-1999 (first entry)
DT NC-contig derived from mardel(10) on chromosome 10q25.2.
DE Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
XX neocentromere; replication; extra-chromosomal element; segregation;
XX cell division; artificial chromosome; gene therapy; mardel(10);
XX human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
OS Homo sapiens.
XX MO9851790-A1.
XX 19-NOV-1998.
XX 13-MAY-1998; 98WO-AU000352.
XX PF 13-MAY-1997; 97AU-00006784.
XX PR 26-AUG-1997; 97AU-00008791.
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX Choo K, Du Sart D, Cancellia MR;
XX WPI; 1999-009773/01.
XX New isolated nucleic acid comprising neocentromere sequences from
XX eukaryotic chromosome - used to produce replicable, segregating
XX artificial chromosomes that can carry large amounts of DNA for gene
XX therapy.
XX Claim 9; Fig 16A; 540bp; English.
XX The present sequence represents the NC-contig derived from a mutated
XX human chromosome 10, 10q25.2 region. The sequence contains an unusual
XX chromosomal marker referred to as mardel(10). The mardel(10) marker is
XX mitotically stable and contains a functional neocentromere at a location
XX regarded as non-centromeric. This neocentromere maps to q25.2 on
XX chromosome 10. The specification describes nucleic acid sequences derived
XX from a eukaryotic chromosome, including a neocentromere or its functional
XX derivative or hybrid, that are able, in a compatible cell, of
XX replicating, acting as extra-chromosomal element and segregating during
XX cell division. The sequences can be used to construct artificial
XX chromosomes for use in gene therapy comprising a replicable, segregating
XX nucleic acid that confers a specific phenotype on cells. Human artificial
XX chromosomes can propagate in human cells and carry large amounts of DNA
XX (e.g. therapeutic genes), and, being extra-chromosomal, they are not
XX transgenic. The artificial chromosomes are also useful for generation of
XX transgenic plants and animals, in production of proteins and to make
XX diagnostic reagents, e.g. for expression of cytokines, receptors and
XX growth factors, or to increase the copy number of a gene in a cell. The
XX constructs may also be used for functional and structural analysis of
XX chromosomes
XX
XX Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;
SQ
XX
XX Query Match 99.9%; Score 1999.4; DB 2; Length 80240;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AAAGCTTCTGAATCAGACAAATGCTTCAACTCTTATACCAACTCTGAGTTGGGC 60
Db 58767 AAAGCTTCTGAATCAGACAAATGCTTCAACTCTTATACCAACTCTGAGTTGGGC 58826
OY 61 GACATGGCTTCCCTTCTAGAGTCTGTGACAGCAATCTTGAATAGTGGCAATTGG 120
Db 58827 GACATGGCTTCCCTTCTAGAGTCTGTGACAGCAATCTTGAATAGTGGCAATTGG 58886
OY 121 GCCGTGATTTTAACTCTTGGTCAAAATTTGTTCTCTAGATCAGGCGCATCAAGCT 180
    |||
    |||
    |||
```

Db 58887 GCCCTGATATTTTAAACCTTGGTCAAATTTTGTTCCTCTAGANTCAGAGCCATCAAGdT 58946  
QY 181 ACAGATGATCTTACAAATGTAACCCCAATGAGCTCAACCTTCTGCTGAGAGC 240  
Db 58947 AAGATGATCTTACAAATGTAACCCCAATGAGCTCAACCTTCTGCTGAGAGC 59006  
QY 241 CCTGACGACCCGCTGAGCCCTTTCATGAGCTTAAAGAGCTCCCTCTGAGAGCACTAC 300  
Db 59007 CCTGACGACCCGCTGAGCCCTTTCATGAGCTTAAAGAGCTCCCTCTGAGAGCACTAC 59066  
QY 301 CACTGAGAGGCCCTTCTTCAACCCCTATCAGAGAGAGTGAAGGCTCATATGCCA 360  
Db 59067 CACTGAGAGGCCCTTCTTCAACCCCTATCAGAGAGAGTGAAGGCTCATATGCCA 59126  
QY 361 AATCCCAACAGAGCTGGGGGTCTGTTTGAAGGGGGATTGAGAGTGAAGCCAGCTG 420  
Db 59127 AATCCCAACAGAGCTGGGGGTCTGTTTGAAGGGGGATTGAGAGTGAAGCCAGCTG 59186  
QY 421 GGGCTTGGGTCAGTGGGGACTTGGAGAACTTTTGTCTAGCTTAAAGATTGTAATG 480  
Db 59187 GGGCTTGGGTCAGTGGGGACTTGGAGAACTTTTGTCTAGCTTAAAGATTGTAATG 59246  
QY 481 CACCAATCAGCACTCTGTGTCTAGCTTAAAGATTGTAATGCAATCAGCACTCTGTA 540  
Db 59247 CACCAATCAGCACTCTGTGTCTAGCTTAAAGATTGTAATGCAATCAGCACTCTGTA 59306  
QY 541 AATAGACCAATCAGAGAGTGGGGGGGTCAAAATGAAGAGTAAATAAGCCAGCC 600  
Db 59307 AATAGACCAATCAGAGAGTGGGGGGGTCAAAATGAAGAGTAAATAAGCCAGCC 59366  
QY 601 GAGCCAGAGTGGCAACCCACTGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 660  
Db 59367 GAGCCAGAGTGGCAACCCACTGGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTTTG 59426  
QY 661 CTCTTCACAATTAATCTTGTCTGCTCATTTCTTTGTCTCAACTTATGAGCTG 720  
Db 59427 CTCTTCACAATTAATCTTGTCTGCTCATTTCTTTGTCTCAACTTATGAGCTG 59486  
QY 721 TAACACTCACTGAGAGGCTGTGGGCTTCACTCTGAGTCAACAGACCAAGCCACT 780  
Db 59487 TAACACTCACTGAGAGGCTGTGGGCTTCACTCTGAGTCAACAGACCAAGCCACT 59546  
QY 781 GGAAGAAACAAGAACTCCGATGTGCTGCTTTAAGCTGTAAACACTGACGAAAdC 840  
Db 59547 GGAAGAAACAAGAACTCCGATGTGCTGCTTTAAGCTGTAAACACTGACGAAAdC 59606  
QY 841 TCTGACGCTTCACTCTGAAAGTCAAGTGAACCAACCAACCAAGAGAAAGAACTCTG 900  
Db 59607 TCTGACGCTTCACTCTGAAAGTCAAGTGAACCAACCAACCAAGAGAAAGAACTCTG 59666  
QY 901 GACACACCTGAATATCTGAAGGAACAACCTCAGACACCACTCTTCAAGGCTGTACA 960  
Db 59667 GACACACCTGAATATCTGAAGGAACAACCTCAGACACCACTCTTCAAGGCTGTACA 59726  
QY 961 CTGACCGCAAGGCTGTGCTTCACTTCTTGAAGTCAAGAACCAAGAAACCCACCGAA 1020  
Db 59727 CTGACCGCAAGGCTGTGCTTCACTTCTTGAAGTCAAGAACCAAGAAACCCACCGAA 59786  
QY 1021 GGAACAAATTCAGACACAGTGAAGAAATCTGATTTTGTGATCTGTGCTTCAAGGTTAC 1080  
Db 59787 GGAACAAATTCAGACACAGTGAAGAAATCTGATTTTGTGATCTGTGCTTCAAGGTTAC 59846  
QY 1081 TCCAGTATGAGATCTCCATTGACGCTTAAAGAAACAAGAAATGTTTGAAGAGCAC 1140  
Db 59847 TCCAGTATGAGATCTCCATTGACGCTTAAAGAAACAAGAAATGTTTGAAGAGCAC 59906  
QY 1141 ATGAGGAATTTGTAATGACACAGGCTTGAATGACATAGGGCACTTCTGATCAACCTTA 1200  
Db 59907 ATGAGGAATTTGTAATGACACAGGCTTGAATGACATAGGGCACTTCTGATCAACCTTA 59966  
QY 1201 GCTGGAAGCAGGGCCAGAAATATATCTTAAGAAACAGTTTTTGTAGACAGTAGTAGT 1260  
Db 59967 GCTGGAAGCAGGGCCAGAAATATATCTTAAGAAACAGTTTTTGTAGACAGTAGTAGT 60026

QY 1261 CTTTGCATCTGAGCATGTGAATTATCAAGCAATTAATTGAAAAATATAGCCAGTGC 1320  
Db 60027 CTTTGCATCTGAGCATGTGAATTATCAAGCAATTAATTGAAAAATATAGCCAGTGC 60086  
QY 1321 GATGCTCATGCTCTTAATCCAGACATTTGGAGGCCAAGGGGTGTGATCAGAGGTGC 1380  
Db 60087 GATGCTCATGCTCTTAATCCAGACATTTGGAGGCCAAGGGGTGTGATCAGAGGTGC 60146  
QY 1381 AGCGCTGAGACACAGCTGGCCCAACATGTAAGAACCCCGTCTTAATAAATACAAA 1440  
Db 60147 AGCGCTGAGACACAGCTGGCCCAACATGTAAGAACCCCGTCTTAATAAATACAAA 60206  
QY 1441 ATTAGCCTGTGTGTGACACGCACTGTATCCAGTACTCAGAGAGCTGAGCGAGG 1500  
Db 60207 ATTAGCCTGTGTGTGACACGCACTGTATCCAGTACTCAGAGAGCTGAGCGAGG 60266  
QY 1501 AATCTCTTGAACCTTGGAGGCGAGAGTGGCAATGAGCCAAAGTCAACCAAGCACTCA 1560  
Db 60267 AATCTCTTGAACCTTGGAGGCGAGAGTGGCAATGAGCCAAAGTCAACCAAGCACTCA 60326  
QY 1561 TCTCGGGTGAACAGACGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAGAAAGAAAT 1620  
Db 60327 TCTCGGGTGAACAGACGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAGAAAGAAAT 60386  
QY 1621 ATATCAAGAAATATGACAGGTAACTTTATTCAACTTACTATGACACAGCAATTACA 1680  
Db 60387 ATATCAAGAAATATGACAGGTAACTTTATTCAACTTACTATGACACAGCAATTACA 60446  
QY 1681 CTAGAGTTTTACATGAGATTAACCTTAATCTTAACAAATAGCCCTATGAAGTCAATGC 1740  
Db 60447 CTAGAGTTTTACATGAGATTAACCTTAATCTTAACAAATAGCCCTATGAAGTCAATGC 60506  
QY 1741 TGTATTAATCTCACTTTAATGAAGAACTGAAGTACAGAAAGTCAAGTAGAGAA 1800  
Db 60507 TGTATTAATCTCACTTTAATGAAGAACTGAAGTACAGAAAGTCAAGTAGAGAA 60566  
QY 1801 TGGCCATGCTTGATCTCAGATTTTGAAGCACTGTTACAGGAATCTGTGTGAGAAAT 1860  
Db 60567 TGGCCATGCTTGATCTCAGATTTTGAAGCACTGTTACAGGAATCTGTGTGAGAAAT 60626  
QY 1861 GCTCTTAACAAGATGTAGTCAAGGGGTGGGAAGTACTGAGTGTGGGAGTTGGGG 1920  
Db 60627 GCTCTTAACAAGATGTAGTCAAGGGGTGGGAAGTACTGAGTGTGGGAGTTGGGG 60686  
QY 1921 ATGGAAGATGATGAAGAAACGCTTGACAGAGAACTGACCTTGGCAACTCTGTGGGA 1980  
Db 60687 ATGGAAGATGATGAAGAAACGCTTGACAGAGAACTGACCTTGGCAACTCTGTGGGA 60746  
QY 1981 CTTTGAAGGTTTGAAGGACT 2001  
Db 60747 CTTTGAAGGTTTGAAGGACT 60767

RESULT 3  
ACN44194  
ID ACN44194 standard; DNA; 275449 BP.  
XX ACN44194;  
XX 18-NOV-2004 (first entry)  
XX Human genomic sequence hCG23995.  
DE  
XX  
XX Cytooblastic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX PN W02003073826-A2.  
XX  
XX 12-SEP-2003.  
PD  
XX  
XX 28-FEB-2003; 2003WO-US006235.  
PF

XX 01-MAR-2002; 2002US-00087192.  
 PR (SAGR-) SAGRES DISCOVERY.  
 PA Morris DW;  
 PI WPI; 2003-328604/31.  
 DR Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
 XX comprises a nucleotide sequence.  
 PT Claim 1; SEQ ID NO 520; Opp; English.  
 XX  
 CC The present invention relates to novel DNA and protein sequences which  
 CC are associated with carcinomas. The sequences are useful for: (i) for  
 CC screening drug candidates; (ii) for screening of bioactive agent capable  
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for  
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the  
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of  
 CC carcinoma including lymphoma. The present sequence is one such CA coding  
 CC sequence. Note: This patent is an equivalent to basic patent  
 CC US2002182586A1, for which no sequence data was published  
 CC  
 SQ Sequence 275449 BP; 80636 A; 53699 C; 54523 G; 86335 T; 0 U; 256 Other;  
 XX  
 Query Match 34.2%; Score 683.4; DB 11; Length 275449;  
 Best Local Similarity 80.3%; Pred. No. 8.3e-154;  
 Matches 931; Conservative 0; Mismatches 101; Indels 127; Gaps 6;  
 QY 1 AAAGGCTTCTGAATGAGACCAATGCTTTCCTTAATACCACTCTGAGTTGGG 60  
 DB 127457 AAAAGCTTCTGATATGAGACCAATGCTTTCCTTAATACCACTCTGAGTTGGG 127516  
 QY 61 GACATGCTTCTGCTCCCTTCTGAGTCTGTGACAGCCATCTTCTAATAGTGGCATTTGG 120  
 DB 127517 AATATGCTTCTTCCCTTCTGAGTCTGTGACAGCCATCTTCTAATAGTGGCATTTGG 127576  
 QY 121 GCCCTGATTTTAACTCTTGTGCAATTTGTTTCTCTAGATGAGGCCATCAAGCT 180  
 DB 127577 GCCCTGATTTTAACTCTTGTGCAATTTGTTTCTCTAGATGAGGCCATCAAGCT 127636  
 QY 181 ACAGATGATCTTCAATGTAACCCCAATGAGTCACTCAACATCTTCTGAGGAGC 240  
 DB 127637 ACAGATGATCTTCAATGTAACCCCAATGAGTCACTCAACATCTTCTGAGGAGC 127696  
 QY 241 CCTGACCGACCGCTGCGCTTTCATGAGCTTAAAGAGTCCCTCTCTGAGGACATAC 300  
 DB 127697 CCTGACCGACCGCTGCGCTTTCATGAGCTTAAAGAGTCCCTCTCTGAGGACATAC 127756  
 QY 301 CACTGACGGGCGCTTCTTCACTCTATCCAGAGAGTCTCAACGCGGTATC-GCC 359  
 DB 127757 AACTTCAGGGCGCTTCTTCACTCTATCCAGAGAGTCTCAACGCGGTATC-GCC 127816  
 QY 360 AATCCCAACAGAGCGTGGGGTTCCTGTTTGAAGGGGGATTTGAGAGGTGAAGCCAGCT 419  
 DB 127817 AATCCCAACAGAGCGTGGGGTTCCTGTTTGAAGGGGGATTTGAGAGGTGAAGCCAGCT 127876  
 QY 420 GGGCTTCTGGGTAGGTGGGACTTGGAGAACTTTTGTCTAGCTAAAGATTTGTAAT 479  
 DB 127877 GAGCTTCTGGGTAGGTGGGACTTGGAGAACTTTTGTCTAGCTAAAGATTTGTAAT 127936  
 QY 480 GCAACCATAGCACTCTGTCTAGCTAAAGATTTGTAATGCAACCATAGCA----- 533  
 DB 127937 GCAACCATAGCACTCTGTCTAGCTAAAGATTTGTAATGCAACCATAGCACTTTGT 127996  
 QY 534 ----- 533

DB 127997 AAAACACACCAATCTGCGCTGTGTCTTAATTAAGTTTGTAAACGACCATCAGCA 128056  
 QY 534 -----CTGTAAAT 544  
 DB 128057 CTCTGTAATAATGACCAATCTGCGCTCTGTAAATGACCAATCAAGTCTGTAAAT 128116  
 QY 545 GGACCAATCAGCAGATGTGGGCGGGTCAATTAAGGAGTAAATAATGCGCACCCGAGC 604  
 DB 128117 GGACCAATCAGCAGATGTGGGCGGGTCAATTAAGGAGTAAATAATGCGCACCCGAGC 128176  
 QY 605 CAGCAGTGGCAACCACTGGGATCCCTTTCACACTGTGGAGCTTTGTTCTTTGCTCT 664  
 DB 128177 TACCAAGGGCAACTGCTGGGATCCCTTTCACACTGTGGAGCTTTGTTCTCTACTCT 128236  
 QY 665 TCACATTAATCTTGTGCTGCTCATTTCTTGTGTCAACTACCTTTAAGCTGTAAAC 724  
 DB 128237 TCACATTAATCTTGTGCTGCTCA--CTCTGGGTTCACTACCTTTAAGCTGTAAAC 128294  
 QY 725 ACTCACTGCGAGGGTCTGTGGCTTCAATTCCTGAAGTCAAC-AGACCAAGAACCACTGGA 783  
 DB 128295 ACTCACTGCGAGGGTCTGTGGCTTCAATTCCTGAAGTCAACAGCACCAACCACTGGA 128354  
 QY 784 AGGACAAAGAACTCCGATGTGCTGCTTTAAGCTGTAACTCACTGCGAAGCTCT 843  
 DB 128355 AGGACAAATTAATCTCC--GCACTACCTTTAAGGCGGTACATCATGTGTGAAGTCT 128410  
 QY 844 GCAGCTTCACTCTGTAAGTCACTGAGACCAACCAACCAAGAAAGAACTGTGGAC 903  
 DB 128411 GCAGCTTCACTCTGTAAGTCACTGAGACCAACCAACCAAGAAAGAACTGTGGAC 128470  
 QY 904 ACACCTGAATATCTGAAGAAAGAACTCCAGACACCAATCTTTCAGGCTGTAACTCTG 963  
 DB 128471 ACATCTGAACATCT---GAACAACTCTGAGACACCAATCTTTAAGAACTGTAACTCTG 128526  
 QY 964 ACCGCAAGGGTCTGTGGCTTCAATTTCTGAAGTCAAGAACCAAGAACCAACCGAAGGA 1023  
 DB 128527 ACTGTAGGGTCTGTGGCTTCAATTTCTGAAGTCAAGAACCAAGAACCAACCGAAGGA 128586  
 QY 1024 ACAATTTCCAGACAGTA 1042  
 DB 128587 ATTAATTCAGACACACTA 128605  
 RESULT 4  
 ADZ12929  
 ID ADZ12929 standard; DNA; 325876 BP.  
 XX  
 AC ADZ12929;  
 XX  
 DT 16-JUN-2005 (first entry)  
 DE Human cancer-associated genomic DNA #38.  
 XX  
 DE Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;  
 KW cytostatic; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO2005031001-A2.  
 PD 07-APR-2005.  
 XX  
 PF 23-SEP-2004; 2004WO-US031617.  
 XX  
 PR 23-SEP-2003; 2003US-00669920.  
 XX  
 PA (CHIR.) CHIRON CORP.  
 XX  
 PI Morris DW, Malandro MS;  
 XX  
 DR WPI; 2005-273395/28.  
 PT Nucleic acid array useful for detecting cancer associated nucleic acid.

PT comprises two or more nucleic acid probes.  
 XX Disclosure; SEQ ID NO 449; 198pp; English.  
 XX The invention relates to a nucleic acid array for detecting a cancer  
 CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.  
 CC The invention also relates to a peptide array comprising two or more  
 CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound  
 CC that binds to a polypeptide, an isolated antibody or its fragment which  
 CC binds to a polypeptide, which is prepared by immunizing a host animal  
 CC with a composition comprising the polypeptide or its antigen binding  
 CC fragment and collecting cells from the host expressing antibodies against  
 CC the antigen or its antigen binding fragment, a composition comprising the  
 CC antibody and a carrier, a method of screening for anticancer activity, a  
 CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
 CC method of treating cancer and a method of inhibiting expression of a CA  
 CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
 CC nucleic acids. The antibody is useful for detecting the presence or  
 CC absence of cancer cells in an individual which involves contacting cells  
 CC from the individual with the antibody and detecting a complex of a CA  
 CC protein from the cancer cells and the antibody, where the detection of  
 CC the complex correlates with the presence of cancer cells in the  
 CC individual. The composition is useful for inhibiting growth of cancer  
 CC cells in an individual or for delivering a therapeutic agent to cancer  
 CC cells in an individual. The invention is also useful for diagnosing  
 CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
 CC a cell. This sequence represents human cancer-associated genomic DNA of  
 CC the invention.

Sequence 325876 BP; 95594 A; 63666 C; 65568 G; 101048 T; 0 U; 0 Other;

Query Match 34.2%; Score 683.4; DB 14; Length 325876;

Best Local Similarity 80.3%; Pred. No. 8, Be-154; Mismatches 931; Conservative 0; Mismatches 101; Indels 127; Gaps 6;

QY 1 AAAAGCTTCTGAATGACAGCAATGCTTCAAACTTTATACCAACTCTGAGATTGGC 60  
 DB 127480 AAAAGCTTCTGAATGACAGCAATGCTTCAAACTTTATACCAACTCTGAGATTGGC 127539  
 QY 61 GAATGCTTCTGCTCCCTTTCTAGGTCTGTGACAGCCATCTTGATATATCGGATTGG 120  
 DB 127540 AATATGCTTCTGCTCCCTTTCTAGGTCTGTGACAGCCATCTTGATATATCGGATTGG 127599  
 QY 121 GCCCTGATTTTAACTTGTGCTTGAATTTGTTTCTCTAGATGAGGCGCATCAAGCT 180  
 DB 127600 GCCCTGATTTTAACTTGTGCTTGAATTTGTTTCTCTAGATGAGGCGCATCAAGCT 127659  
 QY 181 ACAGATGATCTTACCAATGTAACCCCAATGAGCTCAACTAACAATTCTGCTGAGGACT 240  
 DB 127660 ACAGATGATCTTACCAATGTAACCCCAATGAGCTCAACTAACAATTCTGCTGAGGACT 127719  
 QY 241 CTTGACACCGACCGGCTGCTTCAATGCTTAAAGATCCCTCTGAGAGCACTAC 300  
 DB 127720 CTTGACACCGACCGGCTGCTTCAATGCTTAAAGATCCCTCTGAGAGCACTAC 127779  
 QY 301 CACTGACGAGGCGCTTCTTCAACCCCTATCCAGAGAGTACTACGCGGTATC-GCC 359  
 DB 127780 AACTTCAAGGCGCTTCTTCAACCCCTATCCAGAGAGTACTACGCGGTATCACC 127839  
 QY 360 AAATCCCAAGACGAGCTGAGGTGTCTGTTTGAAGGGGGATTGAGAGGTGAACCACT 419  
 DB 127840 AAATCCCAAGACGAGGTGAGGTGTCTGTTTGAAGGGGGATTGAGAGGTGAACCACT 127899  
 QY 420 GGGCTTCTGGGTACGATGGGAGCTTTGAGAACTTTTGTGTAGCTAAAGATTGTAAT 479  
 DB 127900 GAGCTTCTGGGTACGATGGGAGCTTTGAGAACTTTTGTGTAGCTAAAGATTGTAAT 127959  
 QY 480 GCACCAATTCAGCACTGTGTCTAGCTTAAGAGATTGTAAGCAACCAATCGACGCTTTGT 533  
 DB 127960 GCACCAATTCAGCTGTGTCTAGCTTAAGAGATTGTAAGCAACCAATCGACGCTTTGT 128019  
 QY 534 ----- 533

DB 128020 AAAACACACCAATCTGCGCTGTGTCTTAATAAGTTTGTAAACGACCGATCAGCA 128079  
 QY 534 -----CTGTGTAAT 544  
 DB 128080 CTCTGTAATAATGACCAATCTGCGCTGTGTAAATGACCAATCAGTGTCTGTAAAT 128139  
 QY 545 GACCAATTCAGAGATGTTGGGCGGCTCAATTAAGGAGTAAATCTGACCCGAGC 604  
 DB 128140 GACCAATTCAGAGATGTTGGGCGGCTCAATTAAGGAGTAAATCTGACCCGAGC 128199  
 QY 605 CAGCAGTGCACCAACCTCGGGTCCCTTCCACACTGTGAGAGCTTTGTTCTTTGCTCT 664  
 DB 128200 TGGCAGCGGCAACTGCTCGGCTCCCTTCCACAGCTGTGAGAGCTTTGTTCTCACTCT 128259  
 QY 665 TCACATAAATCTGCTGCTGCTCAATCTTTGTGTCACACTACCTTTATGAGCTGTATC 724  
 DB 128260 TCACATAAATCTGCTGCTGCTCA--CTCTGGTTCACACTACCTTTATGAGCTGTATC 128317  
 QY 725 ACTCACTGAGAGGTGTGCTTCAATCTTCTGAAAGTCAAC-AGACCAAGAACCCACTGGA 783  
 DB 128318 ACTCACTGAGAGGTGTGCTTCAATCTTCTGAAAGTCAACAGACCAAGAACCCACTGGA 128377  
 QY 784 AGGAACAAGAACTCCGATGTGCTGCTTAAAGCTGTAAACACTCAGTGAAGCTCT 843  
 DB 128378 AGGAACAAGAACTCCGATGTGCTGCTTAAAGCTGTAAAGCTGTGAAGCTCT 128433  
 QY 844 GCAGCTTCACTCTGTAAGTCAAGTGAAGCAACCAACCAAGAGAGAACTGTGAC 903  
 DB 128434 GCAGCTTCACTCTGTAAGTCAAGTGAAGCAACCAACCAAGAGAGAACTGTGAC 128493  
 QY 904 ACACTGAAATATCTGAAGAGAAACAACTCAGACACCACTTTTCAAGAGCTGTAACTCT 963  
 DB 128494 ACACTGAAATATCTGAAAGAGAAACAACTCAGACACCACTTTTCAAGAGCTGTAACTCT 128549  
 QY 964 ACCGCAAGGTCTGTGCTTCAATCTTGAAGTCAAGAGAAACCAAGAGAGCAACCGGAGGA 1023  
 DB 128550 ACTGTAGAGGTCTGTGCTTCAATCTTGAAGTCAAGAGAAACCAAGAGAGCAACCGGAGGA 128609  
 QY 1024 ACAATTCAGACACAGTA 1042  
 DB 128610 ATTAATTCAGACACACTA 128628  
 RESULT 5  
 AD064676  
 ID AD064676 standard; cDNA; 2066 BP.  
 AC AD064676;  
 AC  
 DT 07-OCT-2004 (first entry)  
 XX  
 DE Novel human cDNA sequence #1837.  
 XX  
 KW ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;  
 KW cytoskeletal; gene therapy; diagnostic marker; morbid state; osteoporosis;  
 KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;  
 KW cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI440981-A2.  
 XX  
 PD 28-JUL-2004.  
 XX  
 XX 21-JAN-2004; 2004EP-00001196.  
 XX  
 XX 21-JAN-2003; 2003JP-00102206.  
 PR 09-MAY-2003; 2003JP-00131392.  
 XX  
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
 PA  
 XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Iseono Y, Nagai K, Irie R;



XX WPI; 2004-535376/52.  
 DR P-PSDB; AD066864.  
 XX  
 PT Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,  
 Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.  
 XX  
 PS Claim 1; SEQ ID NO 1837; 2449bp; English.  
 XX  
 CC The invention relates to 2495 novel polynucleotides (I) and their encoded  
 CC polypeptides, sequences hybridizing to these nucleotides, sequences  
 CC encoding partial polypeptides and sequences having 70% or 90% identity to  
 CC the nucleotide and protein sequences. The nucleotides and polypeptides  
 CC are useful as diagnostic markers or therapeutic target for the diseases  
 CC or morbid states. They are also useful for treating osteoporosis,  
 CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,  
 CC dementia and various cancers. This sequence corresponds to a nucleotide  
 CC sequence of the invention.  
 CC  
 SQ Sequence 2086 BP; 499 A; 556 C; 502 G; 529 T; 0 U; 0 Other;  
 Query Match 32.7%; Score 654.8; DB 12; Length 2086;  
 Best Local Similarity 86.0%; Pred. No. 1.3e-147;  
 Matches 783; Conservative 0; Mismatches 97; Indels 30; Gaps 4;  
 QY 158 TCTAGATGAGGCGCATCAAGCTACAGATGATCTTCAATGTAACCCCAATGAGCTCA 217  
 DB 713 TATAGATGAGGCGCTATCAAGCTACAGATGATCTTCAATGTAACCCCAATGAGCTCA 772  
 QY 218 ACTAACACTTCTGCTGAGGACCCCTGACCCGAGCCGCTGACCCCTTCAATGAGCTCA 277  
 DB 773 ACTAATAA---CTACAGAGAACCCCTGAGAACCCGCTGACCCCTTCAATGAGCTCA 829  
 QY 278 AGCTCCCTCTGAGGAGCACTACCACTGAGGAGCCCTTCTTCAACCCCTATTCAGAGAA 337  
 DB 830 AGTTCCTCTGAGGAGCACTACCACTGAGGAGCCCTTCTTCAACCCCTATTCAGAGAA 889  
 QY 338 AGTAGTACAGGCGGTATC-GCCAAATCCCAAGAGAGCGGTGTCCTGTTGAGAGG 396  
 DB 890 AGTAGTACAGGCGGTATATCAATCCCAATGAGAGCGGTGTCCTGTTGAGAGG 949  
 QY 397 GGGATTGAGAGGTGAAGCAGCTGAGCTTCTGGGTGAGGTGAGGACTTGGAGAACTTTTG 456  
 DB 950 GAGATTGAGAGGTGAAGCAGCTGAGCTTCTGGGTGAGGTGAGGACTTGGAGAACTTTTG 1009  
 QY 457 TGTCTAGTAAAGATTGAATGACCAATGACACTCTGTGTCTAGCTTAAAGATTGT 516  
 DB 1010 TGTCTAGTAAAGATTGAATGACCAATGACACTCTGTGTCTAGCTTAAAGATTGT 1069  
 QY 517 AATATC-----ACCATGACACTCTGTAAATGAGCAA 551  
 DB 1070 AATATGACCAATGACACTCTGTAAATGAGCAAATGAGCAAATGAGCAA 1129  
 QY 552 TCAGCAGATGATGAGGCGGTCAATATGAGAGTAATAATGAGCAACCCGAGCAGCACT 611  
 DB 1130 TCAGTATGATGAGGCGGTCAATATGAGAGTAATAATGAGCAACCCGAGCAGCACT 1189  
 QY 612 GGCACCACTCGGCTCCCTTCCACACTGTGAGAGCTTTGTTCTTTCTTCCAAAT 671  
 DB 1190 GGCACCACTCGGCTCCCTTCCACACTGTGAGAGCTTTGTTCTTCCAAAT 1249  
 QY 672 AATATCTGCTGCTGCTATCTTTGTGTGACACTACCTTTATGAGCTGTAACACTCACT 731  
 DB 1250 AATATCTGCTGCTGCTATCTTTGTGTGACACTACCTTTATGAGCTGTAACACTCACT 1309  
 QY 732 GCGAGGCTGTGAGCTTCAATCTGTAAGTCAAC-AGACCAAGAACCCATGAGAGGAA 790  
 DB 1310 GCGAGGCTGTGAGCTTCAATCTGTAAGTCAACAGACCAAGAACCCATGAGAGGAA 1369  
 QY 791 AAGAACTCCGATGTCTGCTTTTAAAGAGCTGTAACACTCACTGCAAGCTTGCAGCTT 850  
 DB 1370 ATCGACTTGAGACATGCACTTTTAAAGAGCTGTAACACTCACTGCAAGGCTGTGCTT 1429

QY 851 CACTCTGTAAGTCAAGTGAAGCAAAACCCACGAGAGAGAACTCTGACACACTTG 910  
 DB 1430 CACTCTGTAAGTGAAGTGAAGCAAAACCCACGAGAGAGAACTCTGACACACTTG 1489  
 QY 911 AATATCTGAAGAGAACTCTGACACACTCTTTTCAAGAGCTGTAACACTCACTG 970  
 DB 1490 AATATCTGAAGTGAAGTGAAGCAAACTCTGACACCGCACTTTTAAAGTGAACACTCACTG 1549  
 QY 971 GGGTCTGAGCTTCAATCTTGAAGTGAAGCAAAACCCACGAGAGAACTT 1030  
 DB 1550 GGGTCTGAGCTTCAATCTTGAAGTGAAGCAAAACCCACGAGAGAACTT 1609  
 QY 1031 CCAGACACAG 1040  
 DB 1610 CCGACACAG 1619  
 RESULT 6  
 ABD32902  
 ID ABD32902 standard; DNA; 65277 BP.  
 XX AC ABD32902;  
 XX AC  
 XX DT 18-NOV-2004 (first entry)  
 XX DE Human cancer-associated genomic DNA HD18-038.  
 KW Human; ds; cancer-associated protein; gene; cytosolic; cancer;  
 KW Leukemia; lymphoma; CAP.  
 XX OS Homo sapiens.  
 PN WO2004/074320-A2.  
 XX PD 02-SEP-2004.  
 XX PF 17-FEB-2004; 2004WO-US004730.  
 XX PR 14-FEB-2003; 2003US-00367094.  
 XX PR 14-MAR-2003; 2003US-00388838.  
 XX PR 15-APR-2003; 2003US-00417375.  
 XX PR 13-JUN-2003; 2003US-00461862.  
 XX PR 15-SEP-2003; 2003US-00663431.  
 XX PR 15-DEC-2003; 2003US-00737318.  
 XX PA (SAGR-) SAGRES DISCOVERY INC.  
 XX PI Morris DW, Morris DW, Malandro MS;  
 XX PI  
 DR WPI; 2004-652914/63.  
 XX  
 PT New isolated cancer-associated polynucleotides and polypeptides useful  
 PT for diagnosing, preventing or treating cancers, especially lymphoma and  
 PT leukemia, or in screening for agents that modulate cancer.  
 PS Claim 16; seqid 602; 310bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising at least 10  
 CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
 CC in the specification, or its complement. The nucleic acids encode cancer-  
 CC associated proteins. Also included are an expression vector comprising  
 CC the isolated nucleic acid cited above, a host cell comprising the above  
 CC recombinant nucleic acid or expression vector, a microarray for detecting  
 CC a cancer-associated (CA) nucleic acid comprising at least one probe  
 CC comprising at least 10 contiguous nucleotides of any of the above-  
 CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
 CC an open reading frame of a CA sequence selected from any of the 95  
 CC polynucleotide sequences as mentioned in the specification, or its  
 CC complement), an isolated antibody, (or its antigen binding fragment) that  
 CC binds to the above polypeptide, a hybridoma that produces the above  
 CC monoclonal antibody, a pharmaceutical composition comprising the above  
 CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
 CC cells (comprising the antibody cited above, methods for diagnosing cancer



or for detecting the presence or absence of cancer cells in an individual, a method for inhibiting growth of cancer cells in an individual, a method for delivering a therapeutic agent to cancer cells in an individual, an electronic library comprising the above polynucleotide or polypeptide (or their fragments), methods of screening for anticancer activity or for a bioactive agent capable of modulating the activity of a CA protein (CAP), methods for detecting cancer associated with expression of a polypeptide in a test cell sample, a method for treating cancers and a method for inhibiting the expression of CA gene in a cell. The composition and methods are useful for detecting, diagnosing, preventing and treating cancers, especially lymphoma and leukemia. These may also be used in screening for agents that modulate cancer. The present sequence is a human CAP genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 65277 BP; 19651 A; 11706 C; 12664 G; 21256 T; 0 U; 0 Other;

Query Match 30.9%; Score 619; DB 13; Length 65277;  
Best Local Similarity 85.0%; Pred. No. 1.8e-138;  
Matches 771; Conservative 0; Mismatches 110; Indels 26; Gaps 6;

```

OY 1 AAAGCTTCTGAATGAGCAATGCTTTCCTTATATACCACTCTGAGTTGGG 60
    |||||
DB 4321 AAAGCTTTTATATGAGCAATGCTTTCCTTATATACCACTCTGAGTTGGG 4380
OY 61 GACATGCTTCTCCCTTTCTAGTCTCTGAGACCATTTGCTATATGCTGATTTG 120
    |||||
DB 4381 AAGCTGCTTCTCCCTTCTAGTCTCTGAGACCATTTGCTATATGCTGATTTG 4440
OY 121 GCCCTGATTTTAACTCTGCTGCAATTTGTTTCTCTAGAGTGGGCACTCAAGT 180
    |||||
DB 4441 GCTGTGTTTAACTCTGCTGCAATTTGTTTCTCTAGAGTGGGCACTCAAGT 4500
OY 181 ACAGATGATCTTCAATATGATTAACCCCAATGAGCTCAATCAACTTCTGAGAGCC 240
    |||||
DB 4501 ATAGA-GCTTCAATATGAGTAACTTAAATGAGTCAATCAACTTCTGAGAGCC 4559
OY 241 CTTGAGACGACCGGCTGCTTCAATGCTTAAAGCTTCCCTCTGAGAGCACTAC 300
    |||||
DB 4560 CTTGAGTCAATCACTGCTGCTTCTGAGTCTGAGAGTCTTCTGAGAGCACTAC 4619
OY 301 CACTGACAGGCGCTTCTTCAACCCCTTACAGAGAGTCTGAGAGTCTGAGAGCA 360
    |||||
DB 4620 CACTGACAGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4679
OY 361 AATCCCAACAGAGCTGAGGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4739
    |||||
DB 4680 AATCCCAACAGAGCTTCTGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4799
OY 421 GGCTTCTGAGTCTGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480
    |||||
DB 4740 GGCTTCTGAGTCTGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4799
OY 481 CACCAATCAGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 540
    |||||
DB 4800 CACCAATCAGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4843
OY 541 AATGAGCAATGAGAGTCTGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 600
    |||||
DB 4844 AATGAGCAATGAGAGTCTGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4903
OY 601 GAGCAGACAGTGGCAACCTCTGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 660
    |||||
DB 4904 GAGCAGACAGTGGCAACCTCTGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4963
OY 661 CTCTTCAATTAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
    |||||
DB 4964 TTCTTCAATTAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5023
OY 721 TTAACATCACTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
    |||||

```

```

DB 5024 TAA-----CACCAAGAGGCTCTGTGCTTTCATTCAGAGTCAAGACCAAGAACCCAC 5079
OY 780 TGAAGAGCAAAAGAACTCCGATGTGCTGCTTAAAGAGTGTAACTCACTGCGAAG 839
    |||||
DB 5080 CAGAAAGAAACA---AACTCTGATATGACACTTTAAAGGCTGTAACTCACTGCG-AG 5135
OY 840 CTCTGAGCTTCACTCTGAAAGTCAAGTCAAGACCAACCAACCAAGAAAGAAACTCT 899
    |||||
DB 5136 GTCTGCGGCTTCACTTGAAGTCAAGACCAAGAAATCCACAGAAAGATTAATCCC 5195
OY 900 GGACACA 906
    |||||
DB 5196 AAGACACA 5202

```

RESULT 7  
AACT4309/c  
ID AACT4309 standard; cDNA; 1761 BP.

AACT4309;  
02-FEB-2001 (first entry)

Human secreted protein gene 30 SEQ ID NO:40.

Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
antithematic; antiproliferative; cytostatic; cardiant; vasotropic;  
cerebroprotective; neurotropic; neuroprotective; antibacterial; vitruide;  
fungicide; ophthalmological; vulnary; gene therapy; angiogenesis;  
autoimmune disease; hyperproliferative disorder; infection; skin aging;  
wound healing; cardiovascular disorder; cerebrovascular disorder;  
nervous system disorder; food additive; preservative; ss.

Homo sapiens.

WO200057903-A2.

05-OCT-2000.

22-MAR-2000; 2000MO-US007525.

26-MAR-1999; 99US-0126595P.

22-DEC-1999; 99US-0171549P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM, Komatsu G.

WPI; 2000-594630/56.

P-PSDB; AAB39339.

New nucleic acid molecules encoding 48 human secreted proteins for  
diagnosing, preventing, treating or ameliorating medical conditions and  
used as food additives or preservatives.

Claim 1; Page 338; 395pp; English.

The polynucleotide sequences given in AACT4280 to AACT4327 encode the  
human secreted proteins given in AAB39310 to AAB39357. AAB39358 to  
AAB39400 represent human secreted polypeptide sequences and proteins  
homologous to them, which are given in the exemplification of the present  
invention. Human secreted proteins have activities based on the tissues  
and cells the genes are expressed in. Examples of activities include:  
antiarthritic; immunosuppressive; antirheumatic; antiproliferative;  
cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;  
neuroprotective; antibacterial; vitruide; fungicide; ophthalmological;  
and vulnary. The polynucleotides and polypeptides can be used to  
prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
rabbits, goats, horses, cats, dogs, chickens or sheep. They can also  
be used in diagnosing a pathological condition or susceptibility to a  
pathological condition. Disorders which are diagnosed or treated include  
autoimmune diseases, hyperproliferative disorders, cardiovascular  
disorders, cerebrovascular disorders, angiogenesis, nervous system

CC disorders, infections caused by bacteria, viruses and fungi and ocular  
CC disorders. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
CC components. AAC74271 to AAC74279 and AAB39309 represent sequences used in  
CC the exemplification of the present invention

XX Sequence 1761 BP; 409 A; 406 C; 439 G; 507 T; 0 U; 0 Other;

Query Match 29.3%; Score 586; DB 3; Length 1761;  
Best Local Similarity 77.8%; Pred. No. 4,9e-131;  
Matches 826; Conservative 0; Mismatches 115; Indels 121; Gaps 5;

QY 101 TTGCTAAATGCTGCAATTTGGGCGCTGATTTTAACTCTTGCTCAAAATTTGTTCTCT 160  
DB 1761 TTTTCTTTTCTGCTTTGACCTGTATTTTAACTCTTGCTCAAAATTTGTTCTCT 1702  
QY 161 AGATGAGGCGCATCAAGCTACAGATGATCTTACAAATGTAACCCCAATGAGCTCACT 220  
DB 1701 AGAATTGAGGCGCATCAAGCTACAGATGATCTTACAAATGTAACCCCAATGAGCTCACT 1642  
QY 221 AACACCTTCTGCTGAGGAGCCCGTGAACCGACCGCGCTGGCCCTTCAATGGCCCTAAAGAGC 280  
DB 1641 AACACCTTCTGCTGAGGAGCCCGTGAACCGACCGCGCTGGCCCTTCAATGGCCCTAAAGAGT 1583  
QY 281 TCCCTCTCTGAGGAGCACTACACTGACGAGGCGCCCTTCTTACCCCTATCCAGCAGGAAGT 340  
DB 1582 TCCCTCTCTGAGGAGCACTACACTGACGAGGCGCCCTTCTTACCCCTATCCAGCAGGAAGT 1523  
QY 341 AGCTACAGCGGTATC-GCCAAATCCCAACAGACGCTGGGGGTCTCTGTTGAGAGGGGG 399  
DB 1522 AGCTACAGCGGTATCAGCGCAATTCACACACACTGTGTCTGCTTAAAGAGAGG 1463  
QY 400 ATTGAGAGGAGGAGCGAGCTGGGCTTCTGCTGAGGTGGGGAATTGAGAGACTTTGTGT 459  
DB 1462 ATTGAGAGGAGGAGCGAGCTGGGCTTCTGCTGAGGTGGGGAATTGAGAGACTTTGTGT 1403  
QY 460 CTAGCTAAAGATTTGTAATGACCAATGCACTCTGTGTCTAGCTAAAGATTTGTAAT 519  
DB 1402 CTAGCTAAAGATTTGTAATGACCAATGCACTCTGTGTCTAGCTAAAGATTTGTAAT 1343  
QY 520 TGCACCAATCAGACT----- 535  
DB 1342 TGCACCAATCAGACTCTGTAAATAATGACCAATTAGCACTGTGTCTAATAAGATT 1283  
QY 536 ----- 535  
DB 1282 TGTAAAGCAGCAATAGCACTGTGTAAATAATGAGCACTCTGTGTAAATAAGAGG 1223  
QY 536 -----CTGTAAATAATGAGCAATGAGCAATGAGTGTGGGGTCAAAATGAGGAG 584  
DB 1222 CATACGACCCCTGTAAATGAGCAATGAGCAATGAGTGTGGGGTCAAAATGAGGAG 1163  
QY 585 TAAAAATGAGGAGCGGAGCAGAGTGGCAACCACTGGGGTCCCTTTCCACACTGTGG 644  
DB 1162 TAAAAATGAGGAGCGGAGCAGAGTGGCAACCTTTCTAGAGTCCCTTTCCACACTGTGG 1103  
QY 645 AAGCTTTGTTCTTTTCTCTTCAATTAATCTTGTGTCTGCTCAATTTTGTGTCTCA 704  
DB 1102 AAGCTTTTCTTCACTCTTCAATTAATCTTGTGTCTGCTCAATTTTGTGTCTCA 1043  
QY 705 CTACCTTTTGAAGCTTGAACACTCACTGCGAGGCTGTGGCTTCAATCTCTGAAGTCA-A 763  
DB 1042 CTACCTTTTGAAGCTTGAACACTCACTGCGAGGCTGTGGCTTCAATCTCTGAAGTCA 983  
QY 764 CAGACCAAGAACCACTGAGAGGAAGAAAGAAATCTCCGATGTCTGCTTTAAGAGCTGT 823  
DB 982 GAGACCAAGAACCACTGAGAGGAAGAAAGAAATCTCTGAGCGACCACTTTAAGAGCTGT 923

QY 824 AACACTCACTGCGAAGCTCTGACACTTCACTCTCTGAAGTCAAGACCAAAACCCACC 883  
DB 922 AATGCTCACTGTGAAGGTGTGTGCTTCACTCTCTGAAGTCAAGACCAAGCCACC 863  
QY 884 AGAAGAGAAAGAACT---CTGACACACCTGAAATATCTGAAGAAACAAATCTCCAGACAC 940  
DB 862 AGAAGAGAAAGAACTCTCTCCAGACACATCTGAAATATCTGAAGAAACAAATCTCCAGACAC 803  
QY 941 CATCTTTCAGAGCTGTGAACCTACCGCAAGGCTGTGGCTTCAATCTTGAAGTCAAGCA 1000  
DB 802 CATCTTTCAGAGCTGTGAACCTACCTGAGGCTGTGATCTTCAATCTTGAAGTCAAGCG 743  
QY 1001 AGACCAAGAAACCCAGCGAAGAAAGAAACAAATTCAGACAGTA 1042  
DB 742 AGACCAAGAAACCCCTGAGAGAGAACCAATTCAGACATATA 701

RESULT 8  
ID ABD33054 standard; DNA; 22235 BP.  
XX ABD33054;  
AC ABD33054;  
XX 18-NOV-2004 (first entry)  
DT 18-NOV-2004 (first entry)  
DE Human cancer-associated genomic DNA HD23-007.  
KW Human; ds; cancer-associated protein; gene; cytosolic; cancer;  
KW Leukemia; lymphoma; CAP.  
XX Homo sapiens.  
PN WO2004074320-A2.  
XX 02-SEP-2004.  
PD 02-SEP-2004.  
XX 17-FEB-2004; 2004WO-US004730.  
PF 14-FEB-2003; 2003US-00367094.  
PR 14-MAR-2003; 2003US-00388838.  
PR 15-APR-2003; 2003US-00417375.  
PR 13-JUN-2003; 2003US-00461862.  
PR 15-SEP-2003; 2003US-00663431.  
PR 15-DEC-2003; 2003US-00737318.  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX Morris DW, Morris DW, Malandro MS;  
PI MPI; 2004-652914/63.  
DR MPI; 2004-652914/63.  
XX New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and  
PT leukemia, or in screening for agents that modulate cancer.  
XX claim 16; seqid 858; 31opp; English.  
PS The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer

or for detecting the presence or absence of cancer cells in an individual, a method for inhibiting growth of cancer cells in an individual, a method for delivering a therapeutic agent to cancer cells in an individual, an electronic library comprising the above polynucleotide or polypeptide (or their fragments), methods of screening for anticancer activity or for a bioactive agent capable of modulating the activity of a CA protein (CAP), methods for detecting cancer associated with expression of a polypeptide in a test cell sample, a method for treating cancers and a method for inhibiting the expression of CA gene in a cell. The composition and methods are useful for detecting, diagnosing, preventing and treating cancers, especially lymphoma and leukemia. These may also be used in screening for agents that modulate cancer. The present sequence is a human CAP genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences

Sequence 22235 BP; 6027 A; 5133 C; 5249 G; 5826 T; 0 U; 0 Other;

Query Match 24.5%; Score 490.6; DB 13; Length 22235;

Best Local Similarity 66.4%; Pred. No. 1.2e-107;

Matches 869; Conservative 0; Mismatches 379; Indels 60; Gaps 9;

```
QY 323 CCTATTCACAGAGAGTACTACAGGGGTCATCGCAATCCCAAGAGAGCTGGGGT 382
DB 12372 CACCAATCGGACACTGTGTATCTAGCTCAAGTTGTAAACCAATAGACACCTGTGTC 12431
QY 383 TCCGTGTGAGAGGGGGGATTGAGAGTGAAGCCAGCTGGGCTCTGGGAGTGGAGAC 442
DB 12432 TAGCTCAGGGTTTGTGCTGACCAATGACACTCTGCATCTAGCTACTCTGGTGGGAC 12491
QY 443 TTGGAGAACTTTGTG-----TCTAGCTAAGAGATTGTAAATGACCAATCAGCACTC 495
DB 12492 TTGGAGAACTTTGTGAGACCTGTATCTAGCTAATCTGGTGGGAGCGTGAAGAACTT 12551
QY 496 TGTGTCTAGCTAAGAGATTGTAAATGACCAATGACACTCTGTGTA-----543
DB 12552 TGTGTCTAGCTGAGGGATGTGTAACGCAACCAATCAGTCCCTGTCAAAACAGACACTG 12611
QY 544 --TGACCAATTCAGAGAGATGTGGGGGGGTCAATTAAGGAGTAAACAGGACACCG 601
DB 12612 GCTTACCAATTCAGAGAGATGTGGGGGGGTCAATTAAGAGATTAAGAGAGGCTGCCG 12671
QY 602 AGCAGAGAGTGGCAACCACTCGGGTCCCTTCCACACTGTGAAGCTTTGTTCTTTT 661
DB 12672 AGCCAGAGAGTGGCAACCGTCTCGGGTCCGCTTCCACACTGTGAAGCTTTGTTTCC 12731
QY 662 TCTTCAATTAATCTGTGCTGCTGCTCAATCTTTGTGTCCACACTACTTTATGACT 721
DB 12732 TCTTGAATTAATCTGTGCTGCTGCTCACTTTGGGTCCACACTGCTTTATGACT 12791
QY 722 AACACTGACGAGGGGTGTGGCTTCATTCGGAAGTCAACAGCAAGAACCACT 780
DB 12792 GACACTATGAGGAGGTGTGACGTCTACTCTGAAGCCAGCGACCAAGAGCCACC 12851
QY 781 GGAAGAAACAAGAAGTCCGATGTGCTGCTTTAAGAGTGTAACTCACTCGAAGC 840
DB 12852 AGGAGAAAGAACTCACTCAAGCCGCTTTAGAGTGTAACTCACTCGAAGG 12910
QY 841 TCTGAGCTTCACTCTGGAAGTCAAGTGAACCAAAACCAACAGAGAAAGAACTCTG 900
DB 12911 TCTGAGCTTCACTCTG--AGCCAGCGAGACCAACCAAGAAAGAAAGAACTCA 12969
QY 901 GACACACTGAAATCTGAAGAAACAACCTCAAGACACATCTTTAGAGCTGTACA 960
DB 12970 AACACATCCGAACATCAAGAGAAACCTCAAGCGGCAACCTTAAGAGCTGTACA 13028
QY 961 CTCACCGCAAGGCTGTGGCTTCACTTTGAAGTCAAGACCAAGAACCCACCGGAA 1020
DB 13029 CTCACCGCAAGGCTCAAGGCTTCACTTTGAAGTCAAGACCAAGAACCCACCAATT 13088
QY 1021 GGAACAAATTCAGACACAGTGAAGAAATCTGTATTTTGAATCTGTGGCTTCAAGGTTAC 1080
```

```
DB 13089 GCGACACACACCAACCACTTAATGCACTTGTCTCCGTCAAGCCCTAA 13148
QY 1081 TCCAGTCAATGAAGTCTCCATTGACGCTTAAAGAAACAGAGAAAGTTTGGAGAGCAC 1140
DB 13149 GCTTCTGGACAGCGAGAGAGTTCATATGTGGAAACGAGGAGGCGCCAAACAG--- 13205
QY 1141 ATGTGGAAATTTGTATGACCAAGGCTTGAAGTCAATGAGGCAATTTGTATCAACCTTA 1200
DB 13206 -----TTTGTTCATTTTGTGTTTACTTTTAAACCTTATTT 13240
QY 1201 GCTGGAAGAGGGCCAGAAATATAATCTAAGAAAGACAGTTTGTGACAGTACT 1260
DB 13241 ACAATCTTTAATCTAGTAAAGTAAAGAAATTAAGAAATAGTT-----AACGCACT 13294
QY 1261 CTTTGCATCTGACATGTAGATTATCAAGCAATTATTAAGAAATAATTAAGCCAGTCC 1320
DB 13295 GGTTCACATTAATVAGAAAGAACAGCTGTCTTAAAGAAATTTTGGCCGCGCAAT 13354
QY 1321 GATGCTCATGCTGTGTAATCCAGACACTTGGAGAGCCAAAGGGGTGTGATCAGAGTCC 1380
DB 13355 GGTGGCTACACCTGTATATCCAGCAATTTGGAGGCGAGGTTGGCAGATCATAGAGTTC 13414
QY 1381 AGGCTTTCAGACACAGCTGAGCAACATGTAACCCCGTCTTAATAAAATACAAA 1440
DB 13415 AGGAGTTGAGACAGACCTGACATTAACATGCAAAACCGTTTAACTTAAATAACAAA 13474
QY 1441 ATTAGCTGTGTGTGTGAGCAGCATGTGTAATCCAG-TACTCAGAGGCTGAGCAGG 1499
DB 13475 ATTAGCAGGCAATGTGAGCAGCATTAATCTCAGTACTCAGTAGGCTGAGGAGGA 13534
QY 1500 GAATCTTTGAACCTTGGAGAGGAGAGTTCAGAGTGAAGTCAACACAGCACTCC 1559
DB 13535 GAATCACTTGAACCCAGAGAGGAGAGGAGTTCAGAGTGAAGTCAACGCACTGCACTCC 13594
QY 1560 ATCTGGGTGACAGAGCACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1606
DB 13595 AGCTGGGCAACAGAGTGAAGTCTGTCTCAAAAAAAAAATAATAA 13641

RESULT 9
AEA00519/c
ID AEA00519 standard; DNA; 55763 BP.
XX
AC AEA00519;
AC
DT 14-JUL-2005 (first entry)
XX
DE Human epoxide hydrolase (EPHX2) gene SeqID1.
XX
KW SNP detection; haplotype mapping; neutrotropic; neuroprotective;
KW Alzheimers disease; gene; ds.
XX
OS Homo sapiens.
XX
FH Key
FH Location/Qualifiers
FT variation
FT /*tag= a
FT /label= Single nucleotide polymorphism
FT /replace(21845,G)
FT
FT variation
FT /*tag= b
FT /label= Single nucleotide polymorphism
FT /replace(26376,A)
FT
FT variation
FT /*tag= c
FT /label= Single nucleotide polymorphism
FT /replace(26434,C)
FT
FT variation
FT /*tag= d
FT /label= Single nucleotide polymorphism
FT /replace(46768,T)
FT
FT variation
FT /*tag= e
FT /label= Single nucleotide polymorphism
FT /replace(54475,C)
FT
FT variation
FT /*tag= t
FT /label= Single nucleotide polymorphism
```

XX MO2005042706-A2.  
XX  
XX 12-MAY-2005.  
XX  
XX 26-OCT-2004; 2004MO-US035235.  
XX  
XX 28-OCT-2003; 2003US-0515378P.  
XX  
XX (GENA-) GENA1SSANCE PHARM INC.  
XX  
XX Aarsens J, Athanasios M, Brain C, Cohen N, Dain B, Denton RR,  
XX Judson RS, Ozdemir V, Reed CR;  
XX WPI; 2005-366600/37.  
XX  
XX Determining a response marker I or II by determining zero copies or at  
XX least one copy of BPHX2 haplotypes having associations with response to  
XX galantamine, useful for treating patients with Alzheimer's disease.  
XX  
XX Claim 1; SEQ ID NO 1; 142bp; English.  
XX  
XX This invention relates to a novel method of determining whether an  
XX individual has a response marker I or II. The method comprises  
XX determining whether the individual has zero copies or at least one copy  
XX of any of haplotypes 1-26 from 13 BPHX2 haplotypes having associations  
XX with response to galantamine and their corresponding polymorphic sites as  
XX given in the specification as table 1. The invention may be useful for  
XX the development of compounds with a nootropic or neuroprotective activity  
XX acting as epoxide-hydrolase inhibitors. The methods and compositions of  
XX the present invention are useful for treating patients having one of the  
XX BPHX2 haplotypes, predicting the response of an individual to galantamine  
XX and for treating Alzheimer's disease patients based upon their haplotype  
XX profile. The present sequence is that of the human epoxide hydrolase  
XX (BPHX2) gene of which haplotypes associated with galantamine response  
XX were identified for use in the method of the invention.  
XX  
XX Sequence 55763 BP; 13651 A; 13167 C; 13346 G; 15599 T; 0 U; 0 Other;  
XX  
XX Query Match 23.4%; Score 461.8; DB 14; Length 55763;  
XX Best Local Similarity 78.2%; Pred. No. 1.5e-100;  
XX Matches 639; Conservative 0; Mismatches 152; Indels 26; Gaps 6;  
XX  
XX 1 AAGGCTTCTGAATGAGAAATGCTTGAATCTTAACCACTCTGAGTTGGC 60  
XX |||||  
XX 3307 AAGGCTTCTGAACAGCAACACCTTCAAAATCTTAACCACTCTGAGTTGGT 3248  
XX |||||  
XX 61 GACATGGCTTCTCCCTTTCTAGTCTGAGAGCCATCTTGAATGAGATTGG 120  
XX |||||  
XX 3247 GACTTGGCTTCTCCCTTTCTAGTCTGAGAGCCATCTTGAATGAGATTGG 3188  
XX |||||  
XX 121 GCCCTGATTTTAACTCTTGTGCAATTTGTTTCTTGAAGAGCCATCAAGCT 180  
XX |||||  
XX 3187 TCCCTGATTTTAACTCTTCTCTCAAAATTTGTTTCTTCAAGATCAAGCCATCAAGCT 3128  
XX |||||  
XX 181 ACAGATGATTTTAACTTAAATGTAACCCCAATGAGCTCACTTAACAATCTTCTGAGAGATC 240  
XX |||||  
XX 3127 ACAGATGATTTTAACTTAAATGTAACCCCAATGAGCTTAACTCAACAATCTTCAAGAGATC 3068  
XX |||||  
XX 241 CCGAGACGAGCCGCTGAGCCCTTTCAATGAGCTTAAAGAGCTCCCTCTGAGAGACATAC 300  
XX |||||  
XX 3067 GCTAGACTGACCCACCTGAGCCCTTTGACTGAGCTTGAAGAGTTCCCTCTGAGAGACATAC 3008  
XX |||||  
XX 301 CACTGAGAGGCGCCCTTCTTCAACCCCTATCCAGAGAGAAATGATGATCAAGCGGTATC-6TC 359  
XX |||||  
XX 3007 AACGCGAGAGGCGCCCTTCTTCTTGGCCCTATCCAGAGAGAAATGATGATGATCATCACC 2948  
XX |||||  
XX 360 AAATCCCAACAGAGAGCTGGGGGTGCTCTGTTGAGAGGGGATTTGAGAGGTGAGCCAGCT 419  
XX |||||  
XX 2947 AGTTCCAGTAGAGAGGTGGGGGTGCTCTGTTATGAGAGGATTTGAGAGGTGAGCCAGCT 2888  
XX |||||  
XX 420 GGGCTT-CTGGGTCAAGTGGGAGCTTGAAGAACTTTTGTCTAGCTTAAAGATTTG-AA 477  
XX |||||

Db 2887 GGACTTCTAGTGTGAGTGGGAGCTTGGGAACTTCTGTCTTAACAAGAAATTTGTAA 2828  
Qy 478 ATGACCAATACAGACTGTGTCTGCTTAAGATTTGTAATGACCAATACGACTCT 537  
Db 2827 ATGACCAATACAGACTGTGTCTGTAAGACCAATACAGACTTCTTAAGTACCAATCA 2768  
Qy 538 GTAAATGACCAATACGCA-----GGATGTGGCGGGGTCAAA 576  
Db 2767 CGGGAGAGATTGAAAAAAGGCACTGTGATAGACAGAAATGAAATGGAATGGAGAGACAA 2708  
Qy 577 TAAGGAGTAAATAATGAGCCAGCCAGCCAGAGAGTGGCAACCACTCGGGTCCCTTCCA 636  
Db 2707 TAAGGAGTAAATAATGAGCCAGCCAGCCAGAGAGTGGCAACCACTCGGGTCCCTTCCA 2648  
Qy 637 CACTGTGAGAGCTTTTGTCTTGTCTTCTTCAATTAATTTCTGTCTGATTTCTTTG 636  
Db 2647 TGTCTGTGAGAGCTTTTGTCTTGTCTTGTCTTCAATTAATTTCTGTCTGATTTCTTTG 2588  
Qy 697 TGTCCACACTACTCTTATGAGCTGTAACTACTGAGAGGTGTGTGCTTCAATTTCTG 756  
Db 2587 GGTCTGCGCATTTTATGAG- TGTAACTACTGAGAGGTGTGTGCTTCAATTTCTG 2529  
Qy 757 AAGTCAAC-AGACCAAGAACCTGAGAAAGAAACAA 792  
Db 2528 AAGTCAAGAACCAAGAACCAAGAACCAAGAACCA 2492  
XX  
XX RESULT 10  
XX ABZ80229  
XX ID ABZ80229 standard; DNA; 249999 BP.  
XX  
XX ABZ80229;  
XX  
XX 02-JUN-2003 (first entry)  
XX  
XX Human tramdorin gene region genomic DNA SEQ ID NO:26.  
XX  
XX Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;  
XX central nervous system disorder; CNS disorder; multiple sclerosis;  
XX nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; tramd;  
XX tramdorin; human; chromosome 5; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2003016502-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 21-AUG-2002; 2002MO-US026637.  
XX  
XX 21-AUG-2001; 2001US-0313907P.  
XX  
XX 21-AUG-2002; 2002US-00225810.  
XX  
XX (MCLA-) MCLAUGHLIN RES INST.  
XX  
XX Birmingham JR;  
XX  
XX WPI; 2003-278567/27.  
XX  
XX New nucleic acid sequence encoding tramdorin, e.g. mouse strand 1, mouse  
XX strand 2, mouse strand 3, human strand 1, human strand 2, human strand 3, or  
XX rat strand 1, useful for treating CNS, e.g. stroke, multiple sclerosis,  
XX trauma, neuropathic pain.  
XX  
XX Example 6; Fig 9; 177bp; English.  
XX  
XX The present invention describes an isolated nucleic acid sequence  
XX comprising a cDNA sequence encoding mouse tramdorin (strand 2), mouse  
XX strand 3, human strand 1, human strand 2, human strand 3 or rat strand 1, or  
XX the genomic sequence of mouse strand 1 or mouse strand 3. Mouse strand 1 is  
XX located to chromosome 11, whereas human strand 1 is located to chromosome  
XX 5q31-33. The strand sequences have neuroprotective, nootropic, analgesic  
XX and cerebroprotective activities, and can be used in gene therapy. The



AC AAV43155;  
 XX  
 DT 29-DEC-1998 (first entry)  
 XX  
 DE Multiple sclerosis associated retrovirus-1 clone FBd3 DNA sequence.  
 XX  
 KM Multiple sclerosis associated retrovirus-1; MSRV-1; MS; pol gene;  
 KM gag gene; env gene; rheumatoid arthritis-associated virus; ss.  
 XX  
 OS Multiple sclerosis associated retrovirus.  
 XX  
 PN MO9823755-A1.  
 XX  
 PD 04-JUN-1998.  
 XX  
 PF 26-NOV-1997; 97MO-IB001482.  
 XX  
 PR 26-NOV-1996; 96US-00756429.  
 XX  
 PA (INMR) BIO MERIEUX.  
 XX  
 PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G,  
 PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B;  
 XX  
 DR MPI; 1998-322732/28.  
 XX  
 PT New nucleic acid from retroviruses - useful for diagnosis, prevention and  
 PT treatment of, e.g. multiple sclerosis.  
 XX  
 PS Example 7; Fig 13; 286pp; English.  
 XX  
 CC The present sequence represents the multiple sclerosis (MS) associated  
 CC retrovirus-1 (MSRV-1) clone FBd3 DNA sequence. The invention provides  
 CC complete or partial genomic sequences of the MSRV-1 pol gene, gag gene  
 CC and env gene, and polypeptides encoded by these genes. The invention also  
 CC provides antibodies raised against the polypeptides. The genomic  
 CC sequences, polypeptides and antibodies are also claimed useful for  
 CC diagnosing infection by MS and rheumatoid arthritis-associated viruses,  
 CC and also for prevention and treatment of infection with these viruses  
 CC  
 XX  
 SQ Sequence 1859 BP; 635 A; 384 C; 428 G; 412 T; 0 U; 0 Other;

Query Match 22.6%; Score 452.6; DB 2; Length 1859;  
 Best Local Similarity 91.6%; Pred. No. 7.8e-99;  
 Matches 490; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 1 AAAGGCTTGAATGAGCAATGACCTTGAACCTTATATACCACTTGAGTTGGG 60  
 DB 1325 AAAGGCTTGAATGAGCAATGACCTTGAACCTTATATACCACTTGAGTTGGG 1384  
 QY 61 GACATGGCTTCTCCCTTTAGGTCCTGTGACAGCCATTTGCTAATAGTGGCATTTGG 120  
 DB 1385 AACATGGTCTTCCCTTTATAGTCCATGCGCATCTTCTATTACTCCGCTTTGG 1444  
 QY 121 GCCCTGATTTTAACTCTTGCTGCAAAATTTGTTCTCTAGAGATGAGGCGCATCAAGCT 180  
 DB 1445 GCCCTGATTTTAACTCTTGCTGCAAAATTTGTTCTCTAGAGATGAGGCGCATCAAGCT 1504  
 QY 181 ACAGATGATCTTCAATATGTAACCAATGAGCTCAATPAACAATCTTGCTGAGAGAC 240  
 DB 1505 ACAGATGATCTTCAATATGTAACCAATGAGCTCAATPAACAATCTTGCTGAGAGAC 1564  
 QY 241 CCTGAGACGACCGGCTGGCCCTTTCAATGAGCTCAATPAACAATCTTGCTGAGAGAC 300  
 DB 1565 CCTGAGACGACCGGCTGGCCCTTTCAATGAGCTCAATPAACAATCTTGCTGAGAGAC 1624  
 QY 301 CACTGAGAGGCGCTTCTTCAACCCATCCAGAGAGAGATGATACAGGGGTATG-CC 359  
 DB 1625 CACTGAGAGGCGCTTCTTCAACCCATCCAGAGAGAGATGATACAGGGGTATG-CC 1684  
 QY 360 AAATCCCAACAGAGCTGGGGGTCTCTGTTTGGAGGGGGGATTTGAGAGGTGAACGAGCT 419  
 DB 1685 AAATCCCAACAGAGCTGGGGGTCTCTGTTTGGAGGGGGGATTTGAGAGGTGAACGAGCT 1744

QY 420 GGGCTTGGGGTCAGGTCGGGACCTTGAGAACTTTTGTGCTAGCTAAAGATTGTAAAT 479  
 DB 1745 GGACTTCTGGGTGGGGTGGGACTTGAGAACTTTTGTGCTAGCTAAAGATTGTAAAT 1804  
 QY 480 GCACCAATCAGACATCTGTGTCTAGCTAAAGATTGTAAATGACCAATCAGAC 534  
 DB 1805 GCACCAATCAGATCTGTGTCTAGCTAAAGATTGTAAATGACCAATCAGAC 1859

RESULT 13  
 ADB84337  
 ID ADB84337 standard; cDNA; 1859 BP.  
 XX  
 AC ADB84337;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE MSRV-1 env region clone clone FBd3.  
 XX  
 KM MSRV, ss; multiple sclerosis; rheumatoid arthritis; gag; pol;  
 KM reverse transcriptase; ribonuclease H.  
 XX  
 OS Multiple sclerosis associated retrovirus.  
 XX  
 PN US2003039664-A1.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 26-NOV-1997; 97US-00979847.  
 XX  
 PR 26-NOV-1996; 96US-00756429.  
 XX  
 PA (PERR) PERRON H.  
 PA (BESE) BESEME F.  
 PA (BEDJ) BEDIN F.  
 PA (PARA) PARANHOS-BACCALA G.  
 PA (KOMU) KOMURIAN-PRADEL F.  
 PA (JOLI) JOLIVET-REYNAUD C.  
 PA (MAND) MANDRAND B.  
 PA (GARS) GARSON J A.  
 PA (TUKF) TUXE P W.  
 XX  
 PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G,  
 PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Tuxe PW,  
 DR MPI; 2003-512253/48.  
 XX  
 PT New isolated or purified nucleic acid associated with multiple sclerosis  
 PT and/or rheumatoid arthritis, useful for detecting a virus associated with  
 PT multiple sclerosis or rheumatoid arthritis in a biological sample.  
 XX  
 PS Example 7; Fig 13; 193pp; English.  
 XX  
 CC The invention relates to an isolated or purified nucleic acid from a  
 CC virus associated with multiple sclerosis and/or rheumatoid arthritis,  
 CC multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise  
 CC pol, gag or reverse transcriptase genes (or their fragments) encoding the  
 CC proteins or defined peptides (including immunodominant peptides,  
 CC antigenic peptides or conserved motifs) Also included are a process for  
 CC detecting a virus associated with multiple sclerosis or rheumatoid  
 CC arthritis in a biological sample, a nucleic acid probe for the detection  
 CC of a virus associated with multiple sclerosis or rheumatoid arthritis, a  
 CC primer for the amplification by polymerisation of a nucleic acid of a  
 CC viral material associated with multiple sclerosis or rheumatoid  
 CC arthritis, a polypeptide exhibiting an inhibitory activity on the  
 CC proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,  
 CC and an antibody directed against the MSRV-1 virus obtained by  
 CC immunologically reacting a human or animal body or cells with an  
 CC immunogenic agent consisting of the antigenic polypeptide defined above.  
 CC The nucleic acids are useful for detecting a biological sample a virus  
 CC associated with multiple sclerosis or rheumatoid arthritis, or for  
 CC detecting in a biological sample, the presence of or exposure to a virus



CC associated with multiple sclerosis or rheumatoid arthritis. The present  
CC sequence is an MSRV cDNA clone. Note: The SEQ ID numbers for the  
CC sequences as displayed in the main body of the patent do not match the  
CC SEQ ID numbers in the sequence listing. Consequently those sequences  
CC mentioned in the claims may not be the sequences the authors intended to  
CC claim.

XX Sequence 1859 BP; 635 A; 384 C; 428 G; 412 T; 0 U; 0 Other;

XX Query Match 22.6%; Score 452.6; DB 9; Length 1859;

XX Best Local Similarity 91.6%; Pred. No. 7.8e-99; Matches 490; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 1 AAAGCTTCTGAAATGAGACAAATGACCTTTCAACTCTTATACCACTCTGAGTTGGGC 60  
DB 1325 AAAGCTTCTGAAATGAGACAAATGACCTTTCAACTCTTATACCACTCTGAGTTGGGC 1384  
QY 61 GACATGGCTTCTCCCTTTCTAGAGTCTGTGACAGCCATCTTCTAATAGTGGCATTGG 120  
DB 1385 AACATGGTTTCTCCCTTTCTATAGTCCATGCGCATCTTCTATCTCGCTTTGG 1444  
QY 121 GCCCTGATTTTAACTCTTGTGCAATTTGTTTCTCTAGATGAGGCCATCAAGCT 180  
DB 1445 GCCCTGATTTTAACTCTTGTGCAATTTGTTTCTCTAGATGAGGCCATCAAGCT 1504  
QY 181 ACAGATGATCTTCAATGATGACCCCAATGAGCTCACTAACAATCTTGTGAGGAC 240  
DB 1505 ACAGATGATCTTCAATGATGACCCCAATGAGCTCACTAACAATCTTGTGAGGAC 1564  
QY 241 CCTGACCGACCGCTGCGCTTTCAATGAGCTCCCTCTGAGGACACTAC 300  
DB 1565 CCTGACCGACCGCTGCGCTTTCAATGAGCTCCCTCTGAGGACACTAC 1624  
QY 301 CACTGACGAGGCCCCCTTTCACTCTATCCAGCAGGAAGTACTACGCGCTATG -CC 359  
DB 1625 CACTGACGAGGCCCCCTTTCACTCTATCCAGCAGGAAGTACTACGCGCTATG -CC 1684  
QY 360 AAATCCCAACAGACGCTGGGGGTGTCGTTTGAAGGGGGATTGAGAGTGAAGCCAGCT 419  
DB 1685 AAATCCCAACAGACGCTGGGGGTGTCGTTTGAAGGGGGATTGAGAGTGAAGCCAGCT 1744  
QY 420 GGGCTTCTGGGTGAGGTGGGAGCTTGGAGAACTTTTGTCTAGCTAAGGATTGTAAT 479  
DB 1745 GGAATCTTGGGTGGGGGTGGGAGCTTGGAGAACTTTTGTCTAGCTAAGGATTGTAAT 1804  
QY 480 GCACCAATGACACTCTGTGTCTAGCTAAGGATTGTAATGACCAATGACGAC 534  
DB 1805 GCAACCAATGACACTCTGTGTCTAGCTAAGGATTGTAATGACCAATGACGAC 1859

RESULT 14

ADG14783 ID ADG14783 standard; cDNA; 1859 BP.

XX ADG14783;

XX 26-FEB-2004 (first entry)

XX MSRV-1 clone FBd3.

XX ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.

XX Multiple sclerosis associated retrovirus.

XX US2003198647-A1.

XX 23-OCT-2003.

XX 03-APR-2002; 2002US-00114104.

XX 26-NOV-1996; 96US-00756429.

XX 26-NOV-1997; 97US-00979847.

PA (INNR ) BIO MERIEUX.  
XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;  
XX Komrrian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;  
XX WPI; 2004-032461/03.

XX New isolated nucleic acid and their fragments having the pol gene of a  
XX retrovirus, useful for diagnosing, preventing and/or treating multiple  
XX sclerosis and/or rheumatoid arthritis.

XX Example 7; SEQ ID NO 42; 193bp; English.

XX The invention relates to an isolated nucleic acid which comprises the pol  
XX gene of a retrovirus associated with multiple sclerosis or rheumatoid  
XX arthritis. The methods and compositions of the present invention are  
XX useful for diagnosing, preventing and/or treating multiple sclerosis  
XX and/or rheumatoid arthritis. The present sequence is used in the  
XX exemplification of the invention.

XX Sequence 1859 BP; 635 A; 384 C; 428 G; 412 T; 0 U; 0 Other;

XX Query Match 22.6%; Score 452.6; DB 12; Length 1859;

XX Best Local Similarity 91.6%; Pred. No. 7.8e-99; Matches 490; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 1 AAAGCTTCTGAAATGAGACAAATGACCTTTCAACTCTTATACCACTCTGAGTTGGGC 60  
DB 1325 AAAGCTTCTGAAATGAGACAAATGACCTTTCAACTCTTATACCACTCTGAGTTGGGC 1384  
QY 61 GACATGGCTTCTCCCTTTCTAGAGTCTGTGACAGCCATCTTCTAATAGTGGCATTGG 120  
DB 1385 AACATGGTTTCTCCCTTTCTATAGTCCATGCGCATCTTCTATCTCGCTTTGG 1444  
QY 121 GCCCTGATTTTAACTCTTGTGCAATTTGTTTCTCTAGATGAGGCCATCAAGCT 180  
DB 1445 GCCCTGATTTTAACTCTTGTGCAATTTGTTTCTCTAGATGAGGCCATCAAGCT 1504  
QY 181 ACAGATGATCTTCAATGATGACCCCAATGAGCTCACTAACAATCTTGTGAGGAC 240  
DB 1505 ACAGATGATCTTCAATGATGACCCCAATGAGCTCACTAACAATCTTGTGAGGAC 1564  
QY 241 CCTGACCGACCGCTGCGCTTTCAATGAGCTCCCTCTGAGGACACTAC 300  
DB 1565 CCTGACCGACCGCTGCGCTTTCAATGAGCTCCCTCTGAGGACACTAC 1624  
QY 301 CACTGACGAGGCCCCCTTTCACTCTATCCAGCAGGAAGTACTACGCGCTATG -CC 359  
DB 1625 CACTGACGAGGCCCCCTTTCACTCTATCCAGCAGGAAGTACTACGCGCTATG -CC 1684  
QY 360 AAATCCCAACAGACGCTGGGGGTGTCGTTTGAAGGGGGATTGAGAGTGAAGCCAGCT 419  
DB 1685 AAATCCCAACAGACGCTGGGGGTGTCGTTTGAAGGGGGATTGAGAGTGAAGCCAGCT 1744  
QY 420 GGGCTTCTGGGTGAGGTGGGAGCTTGGAGAACTTTTGTCTAGCTAAGGATTGTAAT 479  
DB 1745 GGAATCTTGGGTGGGGGTGGGAGCTTGGAGAACTTTTGTCTAGCTAAGGATTGTAAT 1804  
QY 480 GCACCAATGACACTCTGTGTCTAGCTAAGGATTGTAATGACCAATGACGAC 534  
DB 1805 GCAACCAATGACACTCTGTGTCTAGCTAAGGATTGTAATGACCAATGACGAC 1859

RESULT 15

ADN95683/C ID ADN95683 standard; DNA; 135090 BP.

XX ADN95683;

XX 01-JUL-2004 (first entry)

XX Human BEC/LBC-related gene sequence SegID607.





XX The invention relates to a novel human polynucleotide and the encoded  
CC polypeptide. A polynucleotide of the invention may have a use in gene  
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful  
CC as a primer for synthesizing the polynucleotide or as a probe for  
CC detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are  
CC useful in gene therapy, for developing a diagnostic marker or medicines  
CC for regulating their expression and activity, or as a target of gene  
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides  
CC are useful as pharmaceutical agents. The present sequence represents a  
CC CDNA sequence of the invention.

XX Sequence 4027 BP; 1001 A; 875 C; 982 G; 1169 T; 0 U; 0 Other;

XX Query Match 22.3%; Score 445.4; DB 11; Length 4027;

XX Best Local Similarity 91.2%; Pred. No. 5.5e-97;

XX Matches 517; Conservative 0; Mismatches 46; Indels 4; Gaps 4;

QY 1 AAAGCTTGAATGAGCAATGCTTTC-AAACTTTATACCAACTCTGAGTTGGG 59  
DB 1152 AAAGCTTGAATGAGCAATGCTTTC-AAACTTTATACCAACTCTGAGTTGGG 1093

QY 60 CGACATGCTTCCCTTTCTAGTCTGTGACAGCCATCTTGTAAATGTGCAATTG 119

DB 1092 CAACATGCTTCTCCCTTTCTAGTCTGTGACAGCCATCTTGTAAATGTGCAATTG 1033

QY 120 GGGCCGTATTTTAACTTGTGATCAAAATTTGTTCTCTAGAGTCAGAGCCATCAAG 179

DB 1032 GACCTT-TATTTTAACTTGTGATCAAAATTTGTTCTCTAGAGTCAGAGCCATCAAG 974

QY 180 TACAGATGATCTTCAAAATGTAAACCCAAATGAGCTCAACTCAACTTCTGCTGAGAC 239

DB 973 TACAGATGATCTTCAAAATGTAAACCCAAATGAGCTCAACTCAACTTCTGCTGAGAC 914

QY 240 CCTTGACCGAAGCCGCTGCGCTTTCAATGAGCTTAAAGAGCTCCCTCTGAGAGACATTA 239

DB 913 CCTTGACCGAAGCCGCTGCGCTTTCAATGAGCTTAAAGAGCTCCCTCTGAGAGACATTA 854

QY 300 CCACTGACGAGGAGCCCTTCTTCAACCCCTATCCAGACGAGAACTAGCTACAGCGTCAATCG-C 358

DB 853 CCACTGACGAGGAGCCCTTCTTCAACCCCTATCCAGACGAGAACTAGCTACAGCGTCAATCG-C 794

QY 359 CAAATCCCAACAGCAGCTGAGGATGCTCTGTTTGGAGGGGAGATTGAGAGTGAAGCCAGC 418

DB 793 CAATTCACAGCAGCAGTGGGGTGTCTGTTTGAAGGGGAGATTGAGAGTGAAGCCAGC 724

QY 419 TGGGCTTCTGGGTCAGGTGGGACTTGGAGAACTTTTGTCTAGCTAAAGGATTTGAA 478

DB 733 TGGGCTTCTGGGTCAGGTGGGACTTGGAGAACTTTTGTCTAGCTAAAGGATTTGAA 674

QY 479 TGGACCAATGAGCACTGTGTCTAGCTAAAGGATTTGAAATGACCAATCAGACTCTG 538

DB 673 CACACCAATGAGCACTGTGTCTAGCTAAAGGATTTGAAATGACCAATCAGACTCTG 614

QY 539 T-AAAATGACCAATCAGAGATGTG 564

DB 613 TAAAAAGCAACCAATCAGCGCTCTGTG 587

RESULT 17

ABN95882

ID ABN95882 standard; DNA; 33206 BP.

XX ABN95882;

XX 13-AUG-2002 (first entry)

XX Gene #2380 used to diagnose liver cancer.

DB Gene, liver cancer; ds; hepatocellular carcinoma; hepatotropic;

XX metastatic liver tumour; cytostatic; expression profile; disease state;

KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX

OS Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US030589.

XX 02-OCT-2000; 2000US-0237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer, hepatocellular

PT carcinoma or metastatic liver tumor in a patient, involves detecting the

PT level of expression of two or more genes in a liver tissue sample.

PS Claim 1; SEQ ID NO 2380; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver

CC tumour in a patient, and differentiating metastatic liver cancer from

CC hepatocellular carcinoma in a patient, involving detecting the level of

CC expression of two or more genes represented in ABN93503-ABN97455 in a

CC tissue sample. The method of the invention has hepatotropic, and

CC cytostatic activity. The method is useful for diagnosing and detecting

CC the progression of liver cancer, hepatocellular carcinoma and metastatic

CC liver carcinoma in a patient. The method is useful for identifying

CC expression profiles which serve as useful diagnostic markers as well as

CC markers that can be used to monitor disease states, disease progression,

CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data

CC for this patent did not form part of the printed specification, but was

CC obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences

XX

XX Sequence 33206 BP; 9901 A; 5689 C; 5955 G; 11661 T; 0 U; 0 Other;

QY Query Match 21.8%; Score 436.8; DB 6; Length 33206;

Best Local Similarity 84.1%; Pred. No. 1.3e-94;

Matches 562; Conservative 0; Mismatches 77; Indels 29; Gaps 5;

QY 402 TGAGAGTGAAGCAGCAGCTGAGCTT-CTGGTCAAGTGGGACTTGGAGAACTTTGTGTC 460

DB 500 TGAGAGTGAAGCAGCAGCTGAGCTT-CTGGTCAAGTGGGACTTGGAGAACTTTGTGTC 559

QY 461 TACCTAAAGATTGTAATGACCAATCAGCACTGTGTCTAGCTAAAGATTGTAAT 520

DB 560 TACCTAAAGATTGTAATGACCAATCAGCACTGTGTCTAGCTAAAGATTGTAAT 619

QY 521 GCACCAAT-----CAGCAGCTGTAAATGGAACCAATCAG 555

DB 620 GCACCAATCAGCACTTTGTAATAACAGACCAACAGCACTGTGTAATAAGGCAATCAG 679

QY 556 CAGATGTGGGCGG-GATCAATAAGGAGTAATAAATGCGCACCGGAGCAGCACTGTGTC 614

DB 680 CGGATGTGGGCGGAGAACCAATAAGGAGTAATAAATGCGCACCGGAGCAGCACTGTGTC 739

QY 615 AACCACTGGGATCCCTTTCACACTGTGGAGCTTTGTTCTTGTCTTCAATTA 674

DB 740 AACCTGTGTGTCCTTTCACAGCTGTGGAATCTTGTCTTGTCTTGTCTTCAATTA 799

QY 675 TCTTGTGCTGCTCATTTCTTGTGTCAACATCTTTAAGAGCTGTAACATCAGCGG 734

DB 800 TCTTGTGCTGCTCATTTCTTGTGTGTGCACTTCTTGAAGCTGTGAACATCAGCGG 859

QY 735 AGGATGTGGGCTTCAATCTGTAAGTCAAC-AGACCAAGAACCACTGTAAGAGAAC 793

DB 860 AAGGTCTGTGGCTTCACTCTGTAAGTCAAGAACCTGTAAGAACCACTGTAAGAGAAC 919

QY 794 AACTCCGATGTGCTGCTTGAAGCTGTGAACATCACTGGAAGCTTGCAGCTTCAAC 853

|||||  
DB AACCTCCGAGCAGCCATCTTTAGAGCTGTAACTGTAAGGTCTGCGCTTAC 979  
QY 854 TCCGTAAGTGAAGACCAAAACCCAGAGAAAGAACTTGGACACCTGAAT 913  
DB 980 TCCGTGAAGTGAAGACCAAGAAACCCAGAGAAAGAACTTGGACACCTGAAT 1039  
QY 914 ATCTGAAGAAACAACTCCAGACACCATCTTTCAGAGCTGTAACTGACCGCAAGG 973  
DB 1040 ATCAG-AGGAAACAACTCCGAGACACCATCTTTAAATACTGTATGCTACCGCAAGC 1098  
QY 974 TCTGTGGCTTCATTTCTTGAAGTGAAGACCAAGAAACCCAGAGAAAGAACTTCCA 1033  
DB 1099 TCCGCACTTCATCTTGAAGTGAAGACCAAGAAACCATTTGAAGAAACCAATTCCA 1158  
QY 1034 GACACAGT 1041  
DB 1159 GACACACT 1166

RESULT 18  
ADA03050  
ID ADA03050 standard; DNA; 28670 BP.  
XX ADA03050;  
XX  
XX  
DT 06-NOV-2003 (first entry)  
XX  
XX Human IGGAP1 carcinoma associated gene, SEQ ID NO:1568.  
DE  
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukemia; cytostatic; gene therapy; drug screening;  
XX gene; ds.  
XX Homo sapiens.  
OS  
XX WO2003057146-A2.  
PN  
XX 17-JUL-2003.  
XX  
XX 26-DEC-2002; 2002WO-US041414.  
PF  
XX 26-DEC-2001; 2001US-00035832.  
PR  
XX (SAGR-) SAGRES DISCOVERY.  
XX  
XX Morris DW;  
PI  
XX WPI; 2003-587068/55.  
DR  
XX  
XX  
XX New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX  
XX  
XX Claim 1, SEQ ID NO 1568; 245bp; English.

The invention relates to recombinant carcinoma associated (CA) nucleic acid sequences from mouse and human (ADA01482-ADA03094), and to recombinant carcinoma associated proteins (CAP) encoded by them. The invention also encompasses expression vectors and host cells comprising a CA nucleic acid, a polypeptide (especially an antibody) that specifically binds to the protein, and a biochip comprising CA nucleic acid or fragments thereof. The sequences of the invention were identified using oncogenic retroviruses, which insert into the genome of the host organism at random. Many of these do not carry transduced host oncogenes or pathogenic trans-acting viral genes, meaning that cancer incidence is a direct consequence of the effects of proviral integration into host protooncogenes. The CA nucleic acid sequences can be used to diagnose carcinoma (especially breast cancer, prostate cancer, lymphoma or leukaemia) or a propensity to carcinoma by determination of the sequence of a CA gene, or by determination of CA gene expression in particular tissues. CA nucleic acids, proteins and antibodies are also useful as therapeutic agents and in screening and evaluating drug candidates. The present sequence represents a specifically claimed human CA nucleic acid

CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
SQ Sequence 28670 BP; 7464 A; 6666 C; 6289 G; 8251 T; 0 U; 0 Other;  
Query Match 21.8%; Score 436.6; DB 9; Length 28670;  
Best Local Similarity 88.9%; Pred. No. 1.4e-94;  
Matches 506; Conservative 0; Mismatches 59; Indels 4; Gaps 3;

QY 476 AAATGACCAATCAGACCTGTGTCTAGCTAAAGATGTAAATGACCAATCAGACT 535  
DB 23250 AAACGACCAATCAGACCTGTGTCTAGCTAAAGATGTAAATGACCAATCAGACT 23309  
QY 536 CTGTAAATGACCAATCAGACGAGATGTGGGCGGGCTCAATTAAGGAGTAAATCTGGC 535  
DB 23310 CTGTAAATGACCAATCAGACGAGATGTGGGCGGGCTCAATTAAGGAGTAAATCTGGC 23369  
QY 596 CACCGAGCGAGAG-TGGCAACCCACTGGGGTCCCTTCCACTGTGGAAGCTTGT 654  
DB 23370 CAGCGAAGCGAGAGCAAGCAACCACTGGGGTCCCTTCCACTGTGGAAGCTTGT 23429  
QY 655 CTTTGTCTTCAATTAATCTTGTGCTGCTCATTTCTTGTGTCACACTACTTAT 714  
DB 23430 CTTTGTCTTCAATTAATCTTGTGCTGCTCATTTCTTGTGTCACACTACTTAT 23487  
QY 715 GAGCTGTAACACTCACTGCGAGGGTCTGTGGCTTCAATTCCTGAAGTCAACA-GACCAAGA 773  
DB 23488 GAGCTGTAACACTCAATGAGGGTCTGTGGCTTCAATTCCTGAAGTCAACA 23547  
QY 774 ACCCAATGGAAGAAAGAAAGAACTCCGATGTGCTGCTTAAAGACTGTAACTCACT 833  
DB 23548 ACCCAATGGAAGAAAGAAAGAACTCCGATGTGCTGCTTAAAGACTGTAACTCACT 23607  
QY 834 GCGAAGCTTGCAGCTTCACTCCTGAAGTCAAGTGAAGCAACCAACCAAGGAAGA 893  
DB 23608 GCGAAGTGTGGGCTTCACTCCTGAAGTCAAGTGAAGCAACCAACCAAGGAAGA 23667  
QY 894 AACTCTGACACACCTGAATATCTGAAGAAACCAACTCGACACCACTTTCAAGC 953  
DB 23668 AACTCGGACACATCTGAACATCTGAAGAAACCAACTCGGACACCACTTTAAGAGC 23727  
QY 954 TGTAACTTCAACCGCAAGGCTGTGGCTTCAATTTGAAGTGAAGCAAGCAAGCAAGC 1013  
DB 23728 TGTAACTTCAACCGCAAGGCTGTGGCTTCAATTTGAAGTGAAGCAAGCAAGCATCC 23787  
QY 1014 ACCGGAAGCACAATTTCCAGACACAGTA 1042  
DB 23788 ACCGAAGCAATTAATTTCCAGACACAGTA 23816

RESULT 19  
ADB72788  
ID ADB72788 standard; DNA; 28670 BP.  
XX ADB72788;  
XX  
XX  
DT 04-DEC-2003 (first entry)  
XX  
XX Human IGGAP1 gene.  
DE  
XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
XX  
XX Homo sapiens.  
OS  
XX WO2003008583-A2.  
PN  
XX 30-JAN-2003.  
PD  
XX 26-DEC-2001; 2001WO-US051291.  
XX

```
PR 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX PI Morris DW, Engelhard EK;
XX DR WPI; 2003-239337/23.
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
XX Claim 1; SEQ ID NO 616; 2304bp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
XX nucleotide sequence selected from any of the 660 sequences fully defined
XX in the specification. A polynucleotide of the invention has cytostatic
XX activity, and may have a use in gene therapy, or in a vaccine. The
XX recombinant nucleic acids and polypeptides are useful for treating
XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX sarcomas. The present sequence represents a human gene of the invention.
XX
XX Sequence 28670 BP; 7464 A; 6666 C; 6289 G; 8251 T; 0 U; 0 Other;
SQ
Query Match 21.8%; Score 436.6; DB 10; Length 28670;
Best Local Similarity 88.9%; Pred. No. 1,4e-94;
Matches 506; Conservative 0; Mismatches 59; Indels 4; Gaps 3;
QY 476 AATGACCAATGAGCACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACT 535
DB 23250 AAACGACCAATGAGCACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACT 23309
QY 536 CTGTAAATGACCAATGAGCACTGTGTCTAGCTAAAGATTGTAATGAGCACTGCGC 595
DB 23310 CTGTAAATGAGCACTGTGTCTAGCTAAAGATTGTAATGAGCACTGCGC 23369
QY 596 CACCCGAGCCAGAG-TGGCAACCCACTGCGGTCCCTTCCACACTGTGAACTTTGTT 654
DB 23370 CAGCGAAGCCAGAGCAAGGAGCCCACTCGGTGCTTCCATGCTGTGAACTTTGTT 23429
QY 655 CTTTGTCTTCCCAATTAATCTGTGCTGCTCATTTCTGTGTCCACTGACCTTTAT 714
DB 23430 CTTTGTCTTCCCAATTAATCTGTGCTGCTCATTTCTGTGTCCACTGACCTTTAT 23487
QY 715 GAGCTGTAACACTCACTGCGAGGGTCTGTGCTTCAATTCCTGAAAGTCACCA-GACCA 773
DB 23488 GAGCTGTAACACTCAATGAGGGTCTGTGCTTCAATTCCTGAAAGTCACCA-GACCA 23547
QY 774 ACCCCTGGAAGAACCAAGAACTCCGAGTGTGCTGCTTTAAGAGCTGTAACACTGACT 833
DB 23548 ACCCCTGGAAGAACCAAGAACTCCGAGTGTGCTGCTTTAAGAGCTGTAACACTGACT 23607
QY 834 GCGAAGCTTGCAGCTTCACTCCCTGAGTCAAGTGAAGCCCAAAACCCAGCAAGGAAGA 893
DB 23608 GCGAAGCTTGCAGCTTCACTCCCTGAGTCAAGTGAAGCCCAAAACCCAGCAAGGAAGA 23667
QY 894 AACTCTGGAAGCACTGATATCTGAAGAACCAACTCCAGACACACATCTTTACAGAGC 953
DB 23668 AACTCTGGAAGCACTGATATCTGAAGAACCAACTCCAGACACACATCTTTACAGAGC 23727
QY 954 TGTAAACATGACCGCAAGGGCTGTGTGCTTCAATCTTGAAGTCAAGCAAGCAAGAAACCC 1013
DB 23728 TGTAAACATGACCGTGAAGGGCTGTGTGCTTCAATCTTGAAGTCAAGCAAGCAAGTCC 23787
QY 1014 ACCGGAAGAACCAATTCAGACACAGTA 1042
DB 23788 ACCAAGGAATTAATTCAGACACAGTA 23816
RESULT 20
```

```
ADA66334
ID ADA66334 standard; DNA; 28672 BP.
XX
XX ADA66334;
AC
XX 20-NOV-2003 (first entry)
XX
XX Human IQGAP1 gene genomic DNA sequence.
XX
XX carcinoma-associated gene; CA gene; Rorc gene; MCG15938 gene; BAT1 gene;
XX IQGAP1 gene; IQGAP1 gene; Zfp29 gene; hCG27579 gene; Kcnj9 gene;
XX KCNJ9 gene; Ppp3cc gene; Ppp3cc gene; MCG9110 gene; hCG27579 gene;
XX cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;
XX carcinoma-associated protein; CAP; cytoskeletal; gene therapy; anticancer;
XX vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;
XX prostate cancer; DNA vaccine; animal model; human; ds; IQGAP1.
XX
XX Homo sapiens.
XX
XX OS
XX
XX MO2003053224-A2.
XX
XX 03-JUL-2003.
XX
XX 20-DEC-2002; 2002WO-US041776.
XX
XX 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX PI Morris DW, Engelhard EK;
XX DR WPI; 2003-569168/53.
XX
XX Novel recombinant carcinoma-associated nucleic acid, useful for
XX evaluating the effect of a candidate carcinoma drug, and for diagnosing
XX carcinoma.
XX
XX Claim 1; Page 82-86; 229bp; English.
XX
XX This invention relates to a novel recombinant carcinoma-associated (CA)
XX nucleic acid comprising a fully defined genomic, mRNA or coding sequences
XX of mouse Rorc gene or human RORC gene, mouse MCG15938 or human gene BAT1,
XX mouse IQGAP1 gene or human IQGAP1 gene, mouse Zfp29 gene or human
XX hCG27579 gene, mouse Kcnj9 gene or human KCNJ9 gene, mouse Ppp3cc gene or
XX human Ppp3cc gene, mouse MCG9110 gene or human hCG27579 gene, as given in
XX the specification. CA genes are genes which are preferably expressed in
XX cancer cells, preferably lymphatic, breast, prostate or epithelial cells.
XX A compound which modifies the expression of the CA genes or bind to
XX carcinoma-associated proteins (CAP) may have cytostatic activity and the
XX development of an anticancer vaccine. Therefore the invention may be
XX useful for diagnosis and treatment of carcinomas, especially lymphoma
XX carcinoma, breast cancer and prostate cancer. The CA genes may also be
XX useful as DNA vaccines and for generating animal models of carcinomas.
XX The present sequence is that of the human IQGAP1 gene genomic DNA
XX sequence of the invention.
XX
XX Sequence 28672 BP; 7465 A; 6666 C; 6289 G; 8252 T; 0 U; 0 Other;
SQ
Query Match 21.8%; Score 436.6; DB 9; Length 28672;
Best Local Similarity 88.9%; Pred. No. 1,4e-94;
Matches 506; Conservative 0; Mismatches 59; Indels 4; Gaps 3;
QY 476 AATGACCAATGAGCACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACT 535
DB 23252 AAACGACCAATGAGCACTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACT 23311
QY 536 CTGTAAATGACCAATGAGCACTGTGTCTAGCTAAAGATTGTAATGAGCACTGCGC 595
DB 23312 CTGTAAATGAGCACTGTGTCTAGCTAAAGATTGTAATGAGCACTGCGC 23371
QY 596 CACCCGAGCCAGAG-TGGCAACCCACTGCGGTCCCTTCCACACTGTGAAAGCTTTGTT 654
DB 23371 CACCCGAGCCAGAG-TGGCAACCCACTGCGGTCCCTTCCACACTGTGAAAGCTTTGTT 654
```

Db	23372	CAGGAAAGCAGACGAGGAAACCAACTCGGGTGCCTTCCAGTGTGGAAAGCTTTGT	234311
OY	655	CTTTTGTCTCTTCAACAATPAATCTTGTGCTGTCTCATTCTTTGTGTCACAATACTTTAT	714
Db	23432	CTTTGCGCTTTCACAGTAAATCTTGCTGTGCTCA--CTGTGGGTTCGGACATACTTTAT	23489
OY	715	GAGCTGTAACTCACTGCGAGGGTCTGTGGCTTCAATCTCTGAAGTCAACA-GACCAAGA	773
Db	23490	GAGCTGTAACTCAATGCGAGGGTGTGGGCTGCATTTCTGAAGTCAGCAAGACACAA	23549
OY	774	ACCCACTGGAAGGAACAAAGATCTCCGAGTGTGCTCTTAAAGCTGTAACTCACT	833
Db	23550	ACCCACCGGAGGAACAAACAACCTCAGATGAGCACCTTTAAGCTGTAACTCACT	23609
OY	834	GCGAAGCTCTGCACTTCACTCTGAAGTCAGTGAACCAAAACCAACAGAAAGAA	893
Db	23610	GCGAAGGTGTGCGGCTTCACTCCGAAGTCAGTGAGACACGAACCACTGAAAGGAA	23669
OY	894	AACTCTGGACACACTGAATATCTGAAGGAACAACTCCAGACACACCACTTTCAAGAC	953
Db	23670	AACTCTGGACACACTGTAACTCTGAAGGAACAACTCCGAGACACCACTTTTAAAGC	23729
OY	954	TGTAACTCACTGCGCAAGGCTCTGTGGCTCAATCTTGAAGTCAGCAAGACCAAGACCC	1013
Db	23730	TGTAACTCACTGCGTGAAGGCTCTGTGGCTTCAATCTTGAAGTCAGCAAGACCAAGATCC	23789
OY	1014	ACCGAAGGAACAAATTCAGACACAGTA1042	
Db	23790	ACCAAAAGAAATAATTCAGACGCACTA23818	

## RESULT 21

ADJ25985\_00/c  
WP Sequence split into 17 fragments LOCUS ADJ25985 Accession Adj25985

WP	Fragment Name	Begin	End
WP	ADJ25985_00	1	110000
WP	ADJ25985_01	100001	210000
WP	ADJ25985_02	200001	310000
WP	ADJ25985_03	300001	410000
WP	ADJ25985_04	400001	510000
WP	ADJ25985_05	500001	610000
WP	ADJ25985_06	600001	710000
WP	ADJ25985_07	700001	810000
WP	ADJ25985_08	800001	910000
WP	ADJ25985_09	900001	1010000
WP	ADJ25985_10	1000001	1110000
WP	ADJ25985_11	1100001	1210000
WP	ADJ25985_12	1200001	1310000
WP	ADJ25985_13	1300001	1410000
WP	ADJ25985_14	1400001	1510000
WP	ADJ25985_15	1500001	1610000
WP	ADJ25985_16	1600001	1691139
ID	ADJ25985 standard; DNA; 1691139 BP.		
XX			
XX	ADJ25985;		
XX			
XX	20-MAY-2004 (first entry)		
DT			
XX			
DE	Human phosphodiesterase 4D (PDE4D) genomic DNA.		
XX			
KW	Human; phosphodiesterase 4D; PDE4D; gene; ds; stroke; at-risk haplotype;		
KM	polymorphism; cerebroprotective; vasotropic; chromosome 5q12.		
XX			
OS	Homo sapiens.		
XX			
PN	US2004014099-A1.		
XX			
PD	22-JAN-2004.		
XX			
PF	18-APR-2003; 2003US-00419723.		
XX			
PR	19-MAR-2001; 2001US-00811352.		
PR	04-FEB-2002; 2002US-00067514.		

PR	25-SEP-2002; 2002US-00255120.
XX	
PA	(DECO-) DECODE GENETICS BHF.
XX	
PI	Gretarsdottir S, Jonsdottir S, Reynisdottir ST, Thorleifsson G,
XX	Gulicher JR;
XX	
DR	WPI; 2004-121560/12.
XX	
PT	Diagnosing susceptibility to a stroke in an individual, comprises
XX	screening for an at-risk haplotype in the phosphodiesterase 4D gene.
XX	
PS	Claim 5; SEQ ID NO 1; 417pp; English.
XX	
CC	The invention relates to a method of diagnosing susceptibility to a
CC	stroke in an individual comprising screening for an at-risk haplotype in
CC	the phosphodiesterase 4D (PDE4D) gene that is more frequently present in
CC	an individual susceptible to stroke compared to a healthy individual,
CC	where the presence of the at-risk haplotype is indicative of a
CC	susceptibility to stroke and where the at-risk haplotype increases risk
CC	of stroke significantly. Polymorphisms in the phosphodiesterase 4D
CC	(PDE4D) gene are useful in diagnosing susceptibility to stroke in an
CC	individual, predicting the clinical course and treatment for stroke and
CC	monitoring the effectiveness of treatment. PDE4D nucleic acids,
CC	polypeptides or agents that alter the activity of the polypeptides are
CC	useful for treating or preventing stroke. This sequence represents human
CC	PDE4D genomic DNA of the invention. Note: The sequence data for this
CC	patent is also available in electronic format from USPRO at
CC	seqdata.uspro.gov/sequence.html.
XX	
SEQ	Sequence 1691139 BP; 51168A; 310872C; 321346G; 545521T; 0U; 4440Other;

Query Match	21.8%	Score 436.2;	DB 12;	Length 110000;
Best local similarity	27.1%	Score 443.0;	DB 12;	Length 110000;

Best Local Similarity 87.1%; Pred. NO. 2.0E-94;  
Matches 803; Conservative 1; Mismatches 289; Indels 104; Gaps 9;

Matches 803; Conservative 1; Mismatches 289; Indels 104; Gaps 9;

431 TCAGTGGGCACTTGAGAACTTTGTCTAGCTAAAGATTGTAATGCACCAATCAG 490

Db 99342 TCTGTCGGGAGCTAGAGACTTTTGTGTCCTAGCTCGGGATTGTAAACGCACCAATCAG 99283

491 CACTCTGT-----GTCTAGCTAAGGATTGTAATGCACCAATCAGCACTCTGTAA 542

[illegible]

DB 99282 CACCCCTGTCATTAACAGACCCAATCAGCTCTCTGTAAACAGACCCAATCGGCTCTCTTAAA 99223

543 ATGACCAATCAGCAGATGTGGCGGGTCAAAATAAGGAGTAAAAA CTGGCCACCCGA 602

Db 99222 ATGACCAATCAGCAGACGTGGTGGGCCAGATAAGAGAATAAAGCAGGCTGCCCA 99163

603 GGCAGCAGTGGCAACCCACTGCGGTTCCCTTTCACACTGTTGGAAGCTTTTGTCTCTTTTGCCT 652

1000

Dd 99162 GCCAGCAGTGGGAACCTGCTGAGGTTCCCTTTCATGCTGTGGTAGCTTGTCTTCACT 99103

663 CTTCACATAATCTTGCTGCTCATTTCTTGTCACACACTACTTTATGAGCTGTA 722

Db 99102 CTTGCAATAATCTTGTGTTGCTCACTCTTGGGTCACACACTGCCCTTATGAGCTGTA 99043

[illegible]

123 ATACATCACATGCGAGGCGCAGCGCATTCAATCCGAGAGTCATAC-AGGATCATGATACCCACACG /81

Db 99042 ACACTCACATGAAGTCTGCAGCTTCAACCCCTGAAGCTAGCAAAACCGGAACCGCGG 98983

782 GAAGGAACAAAGAACTCCCGATGTGCTGCCCTTAAAGAGCTGTAACTCACTGCGAAGCT 841

db 98982 GGAGGACGAACTCCGACATG-TGGCCTTAAGAGCTGTAACTCACTGCGAAGT 98924

**0.40** **Cumulative Cumulative Probability Density**

842 CTGCAGCTTCACTCTGAGGTCAGTGAAGCCACAACCCACCAGAGGAAGAACTCTGG 901

Db 98923 TTGCAGCTTCATTCCTG-AGCCAGCGAGACCGTGAACCCACAGAGAAGAACTCCGA 98865

902 ACACACCTGAATATCTGAAGGAACAACTCCAGACACACCATCTTCAGAGCTGTACAC 961

98864 ACACATCCAAACATCAGAAGGAACAACCTCCAGACACGCTGCTTTAAGAACTGTACAT 98805

1000

962 TCACCGCAAGGCTCTGGCTTCACTCTGAAGTCAAGCAAGCCCAAGAACCATCGAAG 1021

```
|||||
Db      TCACGAGAGGGCTCGCGCTTCATCTTGAATCAGTAAGACCAAGAACCCACC----- 98750
QY      1022 GAACAAATTCACAGACAGTAGAATCTGTATTTTGTCTGTGGCTTCCAGGGTTACT 1081
      |||
Db      98749 -----AATCAGAGACACAGCTGCA----- 98731
QY      1082 CCAAGCATTTGAAGTCTCCATTCGACGCTTAAGAAACAGAGATGTTGAGAGACACA 1141
      |||
Db      98730 -----TTTTCACAGCCGGAATTAAGCTTTGGAATCTTGA 98697
QY      1142 TGTGGAAATTTGTTATGACACAGGCTTGAGATGACATAGGCGCATTTCTGATCAAACTTGA 1201
      |||
Db      98696 AGTTACTTTTATTAATGACGCTTTGATGAGCATTAAATCTTATACCTTCTTTAA 98637
QY      1202 CTGGAAGCAGGCGCAGAAATATATCTTAAGAAAGACAGTTTGTAGACAGTAGAGTC 1261
      |||
Db      98636 TGTATTATTAATTTTGAATAATGTTCAAGTACAGTAAACAAAGATGACACAA----- 98583
QY      1262 TTTTCATCTGAGACATGATGATTATCAGCAATTAATTAAGAAATAATATAGCCAGGTGCG 1321
      |||
Db      98582 -CAGTAACCTTTTAAACGATTTCTTTAGATCTTAAAGATAGCTATTTGACCGACGATG 98524
QY      1322 ATGGCTCATGCTGCTGTAATCCAGCACTTTGGAGGCGCAAGGGGTGTGATCAGAGGTCA 1381
      |||
Db      98523 GTGGCTCAGGCTGTATATCCAGACATTTGGAGGCTGAGGCGGACAGATCAGAGGTCA 98464
QY      1382 GCGCTTCAGACCAAGCTGCGCCAACTGATGTAACCCCGCTCTCTAATAAAAAATCA----- 1437
      |||
Db      98463 GGAGATTGAGACCATCTGTGCTAAACAGGTGAACCCCGCTCTAGTAAATAACAAAA 98404
QY      1438 -----AAATTAGCCGTGTGTGTGGTCAGCACTGTAAATCCAG-TACTCAGAG 1487
      |||
Db      98403 AAAAAAAAAAAATTAAGCCCGAGGTGGGCGAGCGCCTTAAGTCTGCTACTTCGGAG 98344
QY      1488 GCTGAGGCAAGGGGAATCTTTGAATTTGGAGGCGAGGTTGCAAGTGAAGCCAAATCACA 1547
      |||
Db      98343 GCTGAGGCAAGGAATATGATGAACCCAGAGGCGAGGCTTGCAATTAAGCCGAGATTGTG 98284
QY      1548 CCAACAGACTTCATCTCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAA 1604
      |||
Db      98283 CCAATTGCACTCAGCCTGGGTGACAGTGAAGTGAAGTCACTCAAAAAAAAAAGATA 98227

RESULT 22
ABD33075/c
ID ABD33075 standard; DNA; 152759 BP.
XX
AC ABD33075;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-002.
XX
KW Human, cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
WP; 2004-499109/47.
```

```
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 10; 182BP; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 152759 BP; 43847 A; 29952 C; 31446 G; 47160 T; 0 U; 354 Other;
Query Match 21.8%; Score 435.8; DB 13; Length 152759;
Best Local Similarity 66.4%; Pred. No. 3,8e-94;
Matches 872; Conservative 0; Mismatches 287; Indels 154; Gaps 11;
QY 431 TCAGGTGGGAGCACTTGGAAGACTTTGTCTGCTGCTAAAGATGTAATGACACCAATCAG 490
      |||
Db 92436 TCAGGTGGGAGAGTGAAGAACTTTGTGCTGCTGAGGATTTGTAATGACACCAATCAG 92377
QY 491 CACTGTGCTAGCTAAAGATTGTAATGACCAATGACCACTGTGTAATGACCA 550
      |||
Db 92376 CACCTGACA-----AAACAGCAATCAGCTCTCTGTAATGACCA 92333
QY 551 ATCAGAGAGATGTGGCGCGGCTCAATTAAGAGATTAATAATGCGCACCCGACGACGAG 610
      |||
Db 92332 ATCAGAGAGATGTGGGTGGGCGCAGATTAAGAGATTAATAAGCGCTGCGTGAACGACGAG 92273
QY 611 TGGCAACCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTCTTTTGTCTTCACA 670
      |||
Db 92272 TGGCAACCTGTGGGGTCCCTTCCACACTGTGGAAGCTTTGTCTTTTGTCTTCACA 92213
QY 671 TAAATCTGTGCTGCTCAATCTTTGTGTCACACTTAAAGACTGTAACTCTAC 730
      |||
Db 92212 TAAATCTGTGCTGCTCACTCTTTGGGTGTCACACTGCTTAAAGCTGTAACTCTAC 92153
QY 731 TGGAGAGGTCTGTGGCTTCAATTCCTGAAGTCAAC-AGACCAGAACCCACTGGAAGAAC 789
      |||
Db 92152 TGGGAAGGTCTGCAAGCTTCACTCTGAAGCCAGACGACCAACCCAGCAAGGAAC 92093
QY 790 AAGAACTCCCAATGTGCTGCTTTAAAGCTGTAACTCACTGGAAGCTGTGACGT 849
      |||
Db 92092 GAACAACTCCAGATGCTCCGCC-TTAAGACTTCAACACTCAGCAAGAGTCCGACGT 92034
QY 850 TCACTGCTGAAGTCAAGAGACCAAAACCCAGAGGAAGAACTCTGGAACACACT 909
      |||
Db 92033 TCACTGCTG-AGCCAGAGAACCAACCACTTACAGAGGAAGAACTCGGAACACTCT 91975
QY 910 GAATATCTGAAGAAACAACTCCAGACACCACTTTCAAGAGCTGTAACTCACTGCA 969
      |||
Db 91974 GAACATCAGAAAGAAACAACTCCAGACACCACTTTAAGAACTGTAACTCACTGCA 91915
QY 970 AGGCTGTGCTTCAATCTTTGAAGTCAAGAACCAAGAACCACTGGAAG 1022
      |||
Db 91914 AGGCTTCATGCTTCAATCTTTGAAGTCAAGAACCAAGAACCACTGGAAG 91855
QY 1023 AACAAATTCAGACAGAGTGAAGAAATCTGTATTTTGTATCTGTGCTT----- 1070
      |||
Db 91854 CTACATGTTTACAGACTGAAATAATCTTTTCTTTTCTTTTCTTTTCTTTTACCTTAAGATG 91795
QY 1071 -----CGAGGTACTCAGTCAATGAAGTCTCAATTGACGCTTAAGAAA 1117
```

Db 91794 TTGAGATCTGAGCTAGCATTCACAAATTGTTATTTGGCTTAAGCAGAGATTATGCAAAAG 91735  
QY 1118 CAGAGATGGTTTGAGAGACATGTGGAAATGTTAT----- 1156  
Db 91734 CTGTCAATTGGCAAAAGCTCATAGCTGTGACATGATATCTCCAGTGGCTATGGGCAT 91675  
QY 1157 ---GACCAAGGCTTGAGATGACATAGGCAATTTTGATCAAACTAGCTGAAGCAGG 1213  
Db 91674 TGGTGACACATGTTGATGTCATGTTAAATGCTTTAAAACTTATGATTTAGCACTG 91615  
QY 1214 CAGGAAATATAA--TCATAAGAACAGTTTTTGTAGACAGTATGCTTTGCATCTG 1271  
Db 91614 AGAAGAAAGTTACTGCACAGAAAGACAAAGACAGACATGGCATGTTATGATGT 91555  
QY 1272 AGACATGTGATATCA----- 1288  
Db 91554 AGTGAACAAATATTTATTAAGACTATTAAACAAAGCAAGTATGTTTCTTGCAAAAGC 91495  
QY 1289 -----AGCAATTATAGAAAAATATAGCCAG 1317  
Db 91494 AACACCAAGCCCTCTATATGAAACCAACATAGAGAAACCTTTAAAGATTAAGCTG 91435  
QY 1318 TGGATGCTCATGCTGTATTCACAGACTTTGGAGGCCAGGCTGTGATCAC--G 1375  
Db 91434 GCTGGTGCTCAACCTGTATCCAGCACTTTGGAGGCTGATGCGATCACCTG 91375  
QY 1376 AGGTCAAGGCTTGAACACAGCTGGCCAAATGTTGAAACCCGCTCTACTTAAATAA 1435  
Db 91374 AGGTCAAGGCTTCAACACAGCTGACCAACATGTGAAACCCCATCTCTACAAAAATA 91315  
QY 1436 CAAAAATTAGCTGCTGCTGAGCAGCATCTGTATCCAG--TACTCAGAGAGCTGAGG 1494  
Db 91314 CAAAAAGTAGCCAGCATGCTGACATGCTGTATCCAGCTACTCAGAGAGCTGAGG 91255  
QY 1495 CAGGGAATCTCTTGAATTTGGAGGACAGAGTTGAGTGCAGATCACACACAGC 1554  
Db 91254 CAGGGAATCTGCTTGAATCTGGAGACAGAGTTGAGTGCAGATTTGCACCACTGC 91195  
QY 1555 ACTCCATCTGGGTGACAGAGCAGACTGCTCTCAAAAAAAAAAAAAA 1607  
Db 91194 ACTCCAGCTGGGAAACAGAGCAAGACTCGTCTCAATTAACAAAAACAAAAA 91142

RESULT 23  
ABX08336\_00/c  
WP Sequence split into 17 fragments LOCUS ABX08336 Accession Abx08336  
WP Fragment Name Begin End  
WP ABX08336\_00 1 110000  
WP ABX08336\_01 100001 210000  
WP ABX08336\_02 200001 310000  
WP ABX08336\_03 300001 410000  
WP ABX08336\_04 400001 510000  
WP ABX08336\_05 500001 610000  
WP ABX08336\_06 600001 710000  
WP ABX08336\_07 700001 810000  
WP ABX08336\_08 800001 910000  
WP ABX08336\_09 900001 1010000  
WP ABX08336\_10 1000001 1110000  
WP ABX08336\_11 1100001 1210000  
WP ABX08336\_12 1200001 1310000  
WP ABX08336\_13 1300001 1410000  
WP ABX08336\_14 1400001 1510000  
WP ABX08336\_15 1500001 1610000  
WP ABX08336\_16 1600001 1691080  
ID ABX08336 standard; DNA; 1691080 BP.  
XX  
XX ABX08336;  
XX  
XX 20-JAN-2003 (first entry)  
XX  
XX Human phosphodiesterase 4D (PDE4D) gene.  
XX

KW Human; Phosphodiesterase 4D; PDE4D; stroke; cerebroprotective; gene; ds;  
KM single nucleotide polymorphism; SNP.  
XX  
OS Homo sapiens.  
FH Key  
FH variation  
FT Location/Qualifiers  
FT 732790  
FT /tag= a  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 735966  
FT /tag= b  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 736226  
FT /tag= c  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 736516  
FT /tag= d  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 850001  
FT /tag= e  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 852776  
FT /tag= f  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 853079  
FT /tag= g  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 853575  
FT /tag= h  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 856468  
FT /tag= i  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 860845  
FT /tag= j  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 870924  
FT /tag= k  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1027267  
FT /tag= l  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1027643  
FT /tag= m  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1027757  
FT /tag= n  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1028146  
FT /tag= o  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1037657  
FT /tag= p  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1044016  
FT /tag= q  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1044045  
FT /tag= r  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1254737  
FT /tag= s  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1254849  
FT /tag= t  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1255763  
FT /tag= u  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1257206  
FT /tag= v  
FT /standard\_name= "Single nucleotide polymorphism"  
FT 1258161  
FT /tag= v  
FT /standard\_name= "Single nucleotide polymorphism"  
FT variation



```
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= x
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ac
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ad
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= aj
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= as
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= at
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= au
```

```
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= av
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= aw
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ax
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ay
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= az
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= ba
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= bb
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= bc
FT /standard_name= "Single nucleotide polymorphism"
FT variation
FT /tag= bd
```

Query Match 21.8%; Score 435.6; DB 6; Length 110000;  
Best Local Similarity 67.1%; Pred. No. 3.8e-94;  
Matches 803; Conservative 0; Mismatches 290; Indels 104; Gaps 9;

```
QY 431 TCAGTGGGACTTGGAGAACTTTGTCTAGCTTAAAGATTGTAAATGACCAATCAG 490
DB 99343 TCTGGTGGGAGGTAGAACTTTGTCTAGCTTGGGATTTGTAAGCAATCAG 99284
QY 491 CACTGTGT-----GTCTAGTAAAGATTGTAAATGCAATCAGCACTCTGTAA 542
DB 99283 CACCTGTCAAAACAGACCAATCAGCTCTGTAAACAGACCAATCGCTCTTTAA 99224
QY 543 ATGACCAATCAGCAGAGATGTGGCGGGTCAAAATAGGAGTAAGGACCCGCA 602
DB 99223 ATGACCAATCAGCAGAGATGTGGCGGGTCAAAATAGGAGTGGCTGCCA 99164
QY 603 GCCAGAGTGGCAACCACTCGGGTCCCTTCCACACTGTGAAAGCTTTGTCTTGTCT 662
DB 99163 GCCAGAGTGGCAACCACTCGGGTCCCTTCCACAGTGTGAGCTTTTCACT 99104
QY 663 CTTCAATTAATCTTGTCTGTCTCACTTCTTGTCTTCACTTCACTTATAGCTGTA 722
DB 99103 CTTGCAATTAATCTTGTCTGTCTCACTTCTTGTCTTCACTTCACTTATAGCTGTA 99044
QY 723 ACACTCACTGCGAGGCTGTGTGCTTCACTTCTTGTCTTCACTTCACTTATAGCTGTA 781
DB 99043 ACACTCACTGCGAGGCTGTGTGCTTCACTTCTTGTCTTCACTTCACTTATAGCTGTA 98984
QY 782 GAAGAAACAAAGACTCCGATGTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 841
DB 98983 GAAGAAACAAAGACTCCGATGTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 98925
QY 842 CTGAGCTTCACTCTGTAAGTCACTGTAAGTCACTTCACTTCACTTCACTTCACTTCACTTCACT 901
DB 98924 TTGAGCTTCACTCTGTAAGTCACTGTAAGTCACTTCACTTCACTTCACTTCACTTCACTTCACT 98866
QY 902 ACACACTGAATATCTGAAGAAACAACTCAACACACCACTTTTCAAGCTGTACAC 961
DB 98865 ACACACTGAATATCTGAAGAAACAACTCAACACACCACTTTTCAAGCTGTACAT 98806
QY 962 TCACCGGAAGGCTGTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1021
DB 98805 TCACTGAGAGGCTGTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 98751
QY 1022 GAACAAATTCAGACAGTAGGAATCTGTATTTTGTATCTGTGCTTCAAGGTTACT 1081
```

```
Db 98750 -----AACTCAGACACAGTGTGA----- 98732
QY 1082 CCAAGTATTGAAGTCTCCATTGACGCTTAAGAGAAACAGAGATGTTTGAGAGACACA 1141
Db 98731 -----TTTTCACAGCCGGAATTGACTTGAATTCCTTA 98698
QY 1142 TGTGGAAATTGTTATATGACACAGGCTTGATGACATAGGAGCAATTTCTGATCAACCTAG 1201
Db 98697 AGTTACTTTTATTTATGACGCTTTGATAGCCATTAATAATCTTATACCTCTTTAA 98638
QY 1202 CTGGAAGCAGGCGCAGAGAAATATATCTAAAGAGACGTTTGTAGACAGTAGTAGTC 1261
Db 98637 TGTATTATTAATTTTGAATTAAGTTCAATTAAGAGAAAACAAGATGACACAAA----- 98584
QY 1262 TTTGTCATCTGACAGCATGTATATCAAGCAATTAATTAGAAAAATATATGCCAGGTGC 1321
Db 98583 -CAGTAACCTTTAAACGATTTTCTTTAGATCTTTAAAGATACGTAATTTCAAGCAGGCATG 98525
QY 1322 ATGAGCTCATGCTGTATATCCAGACCTTTGGAGGCGCAAGGGGTGATCAGAGGTCA 1381
Db 98524 GTGGCTCAGCCCTGTATATCCAGACCTTTGGAGGCTGAGGCGGACGATCAGAGGTCA 98465
QY 1382 GGGCTTCAGACACGAGCTGCGCAACATGTGTAAACCCCTCTCTACTTAAATATACA---- 1437
Db 98464 GGAAGATTGAGACATCTGCTGAACACGAGAAACCCCGCTCTAGTAAATATCAAAAA 98405
QY 1438 -----AAATATGAGCTGTGTGTGTGCGACGATCTGTAAATCCAG-TACTCAGAG 1487
Db 98404 AAAAAAAAAAAATTTAGCCCGAGTGGGGCAGCGCCTGTAGTCCGCTACTCGCGAG 98345
QY 1488 GCTGAGGCGAGGGGAATCTTTGAACTTTGGAGGCGAGGCTTGAAGTGAAGGCAAGTACA 1547
Db 98344 GCTGAGGCGAGGAATATGATGATGAACCCAGAGGCGGAGCTTGAATTAAGCCGAGTTGTG 98285
QY 1548 CCACAGCACTCCATCTCGGTGACAGAGCAGACTGTGTCTCAAAAAAAAAAAAAA 1604
Db 98284 CCATTGCATCTCAGCCTGGGTGACAGTGTAGACTCCATCTCAAAAAAAAAAAAAAGAT 98228

RESULT 24
ADN97989_00/c
WP Sequence split into 17 fragments LOCUS ADN97989 Accession Adn97989
WP Fragment Name Begin End
WP ADN97989_00 1 110000
WP ADN97989_01 100001 210000
WP ADN97989_02 200001 310000
WP ADN97989_03 300001 410000
WP ADN97989_04 400001 510000
WP ADN97989_05 500001 610000
WP ADN97989_06 600001 710000
WP ADN97989_07 700001 810000
WP ADN97989_08 800001 910000
WP ADN97989_09 900001 1010000
WP ADN97989_10 1000001 1110000
WP ADN97989_11 1100001 1210000
WP ADN97989_12 1200001 1310000
WP ADN97989_13 1300001 1410000
WP ADN97989_14 1400001 1510000
WP ADN97989_15 1500001 1610000
WP ADN97989_16 1600001 1691138
ID ADN97989 standard; DNA; 1691138 BP.
AC ADN97989;
XX
XX 01-JUL-2004 (first entry)
XX
XX Human phosphodiesterase 4D genomic sequence.
XX
XX cerebroprotective; vasotrophic; PDE4D modulator;
KW phosphodiesterase 4D modulator; phosphodiesterase-4D-Modulator; stroke;
KM
XX
```

```
OS Homo sapiens.
PN
XX NC_02004028341-A2.
XX
PD 08-APR-2004.
XX
PF 25-SEP-2003; 2003MO-US029906.
XX
PR 25-SEP-2002; 2003US-00255120.
PR 18-APR-2003; 2003US-00419723.
PR 27-AUG-2003; 2003US-00650120.
XX
PA (DECO-) DECODE GENETICS EHF.
PI
PI Gtetradsdotlr S, Reynisdottir ST, Thorleifsson G, Gulcher JR;
XX
XX WPI; 2004-315987/29.
XX
DR
PT Assessing or diagnosing susceptibility to stroke, comprises determining
PT phosphodiesterase 4D isoform expression levels in an individual compared
PT to control, a difference in isoform expression indicates susceptibility
PT to stroke.
XX
XX Example 1; SEQ ID NO 105; 521bp; English.
XX
PS
CC The invention relates to a method of assessing susceptibility to stroke
CC in an individual by determining phosphodiesterase 4D (PDE4D) isoform
CC expression levels in the individual compared to control, where a
CC difference in isoform expression indicates susceptibility to stroke. The
CC method is useful for assessing susceptibility to stroke (claimed), for
CC predicting those at higher risk of developing a stroke, for specifically
CC identifying a rate-limiting pathway involved in stroke, for predicting an
CC individual's response to a particular drug, even drugs that do not act on
CC PDE4D or its pathway. Modulators of PDE4D isoform are useful for
CC preventing the occurrence of stroke in an individual. PDE4D polypeptides
CC are useful as molecular markers, to raise antibodies or to elicit
CC an immune response, or as markers for cells or tissues in which the
CC corresponding polypeptide is preferentially expressed during tissue
CC differentiation or in a diseased state. This sequence corresponds to the
CC PDE4D genomic sequence.
XX
XX
SQ Sequence 1691138 BP; 511166A; 310874C; 323132G; 545522T; 0U; 440Other;
Query Match 21.8%; Score 435.6; DB 12; Length 110000;
Best Local Similarity 67.1%; Pred. No. 3.8e-94;
Matches 803; Conservative 0; Mismatches 290; Indels 104; Gaps 9;
QY 431 TCAGGCGGGAAGCTTTGAGAACTTTGTGTAGCTAAAGGATGTAAATGACCAATCAG 490
Db 99343 TCTGGTGGGAGGATGAGAACTTTGTGTAGCTGCGGATTTGTAACGACCAATCAG 99284
QY 491 CACTCTGT-----GTCTAGCTAAAGATTTGTAATGACCAATCAGCACTCTGTAA 542
Db 99283 CACCTGTCAAAACAGACCAATCAGCTCTGTAAACAGACCAATCGGCTCTTTAA 99224
QY 543 ATGACCAATCAGCAGAGATGTGGCGGGGTCAATAGGAGTAATAATGCGCACCGCA 602
Db 99223 ATGACCAATCAGCAGAGATGTGGCGGGGTCAATAGGAGTAATAATGAGCGCTGCCAA 99164
QY 603 GCAGAGGTGGCAACCACTCGGGGTCCCTTCCACATGTGAAGTTGTTCTTTGCT 662
Db 99163 GCAGAGGTGGCAACCTCGAGAGTCCCTTCAAGCTGTGAGTGTGTTCTTCACT 99104
QY 663 CTTCACATTAATCTTGTCTGCTCTCATTTTGTGTCCACATACCTTATGAGCTGTA 722
Db 99103 CTTCGATTAATCTTGTGTGCTCACTTTGGGTCCACATGCTTTATGAGCTGTA 99044
QY 723 AACTCAGTGCAGAGGTCTGTGCTTCACTTCTGAAGTAAAC-AAGACAGAACCACTG 781
Db 99043 AACTCAGTGCAGAGGTCTGTGCTTCACTTCTGAAGTAAACCGCGAACCAGCG 98984
QY 782 GAAGAAACAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACACTCAGTGCAGAGCT 841
```

```
Dh 98983 GGAGGACGACGAACACTCCGACATG-TGGCTTAAAGCTGTAAACACTCACTCGGAAGGT 98925
Qy 842 CTGCGCTTCACTCCCTGAAGTCAGTGAGCCAAACCCACGAGAGGAAGAACTCTGG 901
Dh 98924 TTGCGCTTCATCTCTG-AGCCAGGAGACCGTGAACCCACGAGAGGAAGAACTCTCGA 98866
Qy 902 ACACACCTGAATATCTGAAGAAACAACCTCAGACACACACTCTTTCAGAGCTGTAAAC 961
Dh 98865 AACATCCAAACATCCAGAAAGAAACAACATCCAGACACGCTGCTTAAAGATGTAACT 98806
Qy 962 TCACCGCAAGGCTGTGCTTCACTTCTGAAGTCAAGAACCAAGAACCCGGAAG 1021
Dh 98805 TCACGTAGAGGGCTCTGCGCTTCATCTTGAATCAGTAAAGCAAGAACCCGAC 98751
Qy 1022 GAACAAATTCACACACAGTAGAAGAAATCTGTATTTTGTGCTTCCAGGGTTACT 1081
Dh 98750 -----AACTCAGACACAGGTGA----- 98732
Qy 1082 CCAGTCATTGAAGTCTCAATTGACGCTTAAAGAAACAGAGAAATGTTTGAAGACACA 1141
Dh 98731 -----TTTTCACAGCCGGAATTTGACTTGAAGATCTTAA 98698
Qy 1142 TGTGGAAATTTTATGAGACGAGCTTGAGATGCAATAGGCACTTCTGATCAACCTAG 1201
Dh 98697 AGTTACTTTTATGAGCTTTTGAATGACCAATAAAATCTTATACCTTCTTAA 98638
Qy 1202 CTGGAAGCAGGCGCAGAAATATATCTAAGAAAGAGTTTGTAGACATGATAGTC 1261
Dh 98637 TGTATTTTAAATTTTGAATTAAGTTCACTAACAAGAAACAGATGACACAA----- 98584
Qy 1262 TTTGCATCTGAGACATGTATATCAAGCAATTAATTAAGAAATATATAGCAGGTGC 1321
Dh 98583 -CAGTAACTTTAAAGCATTTTCTTATGATCTTAAAGATTAACGATTTTCAGCCAGGATG 98525
Qy 1322 ATGCTCATGCTGTATATCCAGACCTTGGAGGCGCAGGCTGTGATCAGAGTCA 1381
Dh 98524 GTGGCTACGCTGTATATCCAGACTTTGGAGAGCTGAGGCGGAGATCAGAGGTCA 98465
Qy 1382 GCGCTTGAAGCCAGCTGGCCAAAGTGTGAACCCCGTCTCTACTAATAATCA----- 1437
Dh 98464 GGAGATTGAGACCACTCTGCTAACAAGTGAACCCCGCTTGAATAATAACAATA 98405
Qy 1438 -----AAATTAAGCTGTGTGTGAGCAGCATCTGTATCCAG-TACTCAGAG 1487
Dh 98404 AAAAAAAAAAAAAAAAAATTAAGCTGGAGAGTGGGGGAGCGCTGTAGTCTGTCTGCGGAG 98345
Qy 1488 GCTGAGCAGGAGGAATCTTTGAACCTTGGAGGAGAGGTTGCAGTAGCCAGATCA 1547
Dh 98344 GCTGAGCAGGAGGAATGACATGAACCCAGAGGCGGAGCTTGCATTAAGCCGAGATTGTG 98285
Qy 1548 CCACAGCATCTCTCTGGGTGACAGACCGAGACTTGTCTCAAAAAAAAAAAAAA 1604
Dh 98284 CCAATTGACCTCCAGCTGGGTGACAGTGTGAGACTTCATCAAAAAAAAAAAGATA 98228
```

## RESULT 25

AD050281\_00/c

```
WP Sequence split into 17 fragments LOCUS AD050281 Accession Ad050281
WP Fragment Name Begin End
WP AD050281_00 1 110000
WP AD050281_01 100001 210000
WP AD050281_02 200001 310000
WP AD050281_03 300001 410000
WP AD050281_04 400001 510000
WP AD050281_05 500001 610000
WP AD050281_06 600001 710000
WP AD050281_07 700001 810000
WP AD050281_08 800001 910000
WP AD050281_09 900001 1010000
WP AD050281_10 1000001 1110000
WP AD050281_11 1100001 1210000
WP AD050281_12 1200001 1310000
WP AD050281_13 1300001 1410000
```

```
WP AD050281_14 1400001 1510000
WP AD050281_15 1500001 1610000
WP AD050281_16 1600001 1691134
ID AD050281 Standard; DNA; 1691134 BP.
XX
AC AD050281;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human phosphodiesterase 4D (PDE4D) gene.
XX
KW ds; stroke; phosphodiesterase 4D; PDE4D; human; gene; SNP;
XX single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
FH Key
FH variation
FT replace(135641,a)
FT /tag= a
FT /note= "Single nucleotide polymorphism"
FT replace(732790,c)
FT /tag= b
FT /note= "Single nucleotide polymorphism"
FT replace(73586,c)
FT /tag= c
FT /note= "Single nucleotide polymorphism"
FT replace(736516,c)
FT /tag= d
FT /note= "Single nucleotide polymorphism"
FT replace(850001,c)
FT /tag= e
FT /note= "Single nucleotide polymorphism"
FT replace(852776,a)
FT /tag= f
FT /note= "Single nucleotide polymorphism"
FT replace(853575,c)
FT /tag= g
FT /note= "Single nucleotide polymorphism"
FT replace(856468,c)
FT /tag= h
FT /note= "Single nucleotide polymorphism"
FT replace(870924,c)
FT /tag= i
FT /note= "Single nucleotide polymorphism"
FT replace(1027643,c)
FT /tag= j
FT /note= "Single nucleotide polymorphism"
FT replace(104016,c)
FT /tag= k
FT /note= "Single nucleotide polymorphism"
FT replace(1254849,c)
FT /tag= l
FT /note= "Single nucleotide polymorphism"
FT replace(1258161,c)
FT /tag= m
FT /note= "Single nucleotide polymorphism"
FT replace(1268007,g)
FT /tag= n
FT /note= "Single nucleotide polymorphism"
FT replace(1268187,c)
FT /tag= o
FT /note= "Single nucleotide polymorphism"
FT replace(1268553,c)
FT /tag= p
FT /note= "Single nucleotide polymorphism"
FT replace(1272669,c)
FT /tag= q
FT /note= "Single nucleotide polymorphism"
FT replace(1272910,a)
FT /tag= r
FT /note= "Single nucleotide polymorphism"
FT replace(1273220,c)
FT /tag= s
```

FT /note= "Single nucleotide polymorphism"  
FT replace(1273240,c)  
FT /\*tag= t  
FT /note= "Single nucleotide polymorphism"  
FT replace(1345286,c)  
FT /\*tag= u  
FT /note= "Single nucleotide polymorphism"  
FT replace(1346112,c)  
FT /\*tag= v  
FT /note= "Single nucleotide polymorphism"  
FT replace(1354377,c)  
FT /\*tag= w  
FT /note= "Single nucleotide polymorphism"  
FT replace(1354554,a)  
FT /\*tag= x  
FT /note= "Single nucleotide polymorphism"  
FT replace(1354675,c)  
FT /\*tag= y  
FT /note= "Single nucleotide polymorphism"  
FT replace(1357081,a)  
FT /\*tag= z  
FT /note= "Single nucleotide polymorphism"  
FT replace(1363021,c)  
FT /\*tag= aa  
FT /note= "Single nucleotide polymorphism"  
FT replace(1373349,c)  
FT /\*tag= ab  
FT /note= "Single nucleotide polymorphism"  
FT replace(1373384,g)  
FT /\*tag= ac  
FT /note= "Single nucleotide polymorphism"  
FT replace(1376149,g)  
FT /\*tag= ad  
FT /note= "Single nucleotide polymorphism"  
FT replace(1385445,c)  
FT /\*tag= ae  
FT /note= "Single nucleotide polymorphism"  
FT replace(1409210,a)  
FT /\*tag= af  
FT /note= "Single nucleotide polymorphism"  
FT replace(1414804,t)  
FT /\*tag= ag  
FT /note= "Single nucleotide polymorphism"  
FT replace(1428284,c)  
FT /\*tag= ah  
FT /note= "Single nucleotide polymorphism"  
FT replace(1431800,t)  
FT /\*tag= ai  
FT /note= "Single nucleotide polymorphism"  
FT replace(1449904,a)  
FT /\*tag= aj  
FT /note= "Single nucleotide polymorphism"  
FT replace(1574301,c)  
FT /\*tag= ak  
FT /note= "Single nucleotide polymorphism"  
FT replace(1574615,c)  
FT /\*tag= al  
FT /note= "Single nucleotide polymorphism"  
FT replace(1575634,a)  
FT /\*tag= am  
FT /note= "Single nucleotide polymorphism"  
FT replace(1580088,g)  
FT /\*tag= an  
FT /note= "Single nucleotide polymorphism"  
FT replace(1581078,a)  
FT /\*tag= ao  
FT /note= "Single nucleotide polymorphism"  
FT replace(1582418,t)  
FT /\*tag= ap  
FT /note= "Single nucleotide polymorphism"  
FT replace(1590837,a)  
FT /\*tag= aq  
FT /note= "Single nucleotide polymorphism"

FT variation replace(1590936,a)  
FT /\*tag= ar  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1591011,g)  
FT /\*tag= as  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1591047,t)  
FT /\*tag= at  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1591306,a)  
FT /\*tag= au  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1591583,c)  
FT /\*tag= av  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1594994,g)  
FT /\*tag= aw  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1636902,c)  
FT /\*tag= ax  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1638550,c)  
FT /\*tag= ay  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1640663,c)  
FT /\*tag= az  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1641954,t)  
FT /\*tag= ba  
FT /note= "Single nucleotide polymorphism"  
FT variation replace(1655748,a)  
FT /\*tag= bb  
FT /note= "Single nucleotide polymorphism"  
XX US2004091865-A1.  
XX  
XX  
XX 13-MAY-2004.  
XX  
Query Match 21.8%; Score 435.6; DB 12; Length 110000;  
Best Local Similarity 67.1%; Pred. No. 3.8e-94;  
Matches 803; Conservative 0; Mismatches 290; Indels 104; Gaps 9;  
QY 431 TCAGTGGGAGCACTTGAGAACTTTGTGCTGCTAAAGATGTAAATGCAACCAATCG 490  
DB 99343 TCTGTGGGGAGGTAGAGAACTTTGTGCTGCTGCGGATGTAAAGCAACCAATCG 99284  
QY 491 CACTCTGT-----GTCTAGCTAAAGATTGTAAATGCAACCAATCAGCACTGTAA 542  
DB 99283 CACCTGTCAAAACAGACCAATCAGCTCTCTGTAAACAGACCAATCGCTCTTTAA 99224  
QY 543 ATGACCAATTCAGCAGATGTGGCGGGGTCAATTAAGGAGTAAATACTGGCCACCGA 602  
DB 99223 ATGACCAATTCAGCAGAGTGGGTGGGCGAGATTAAGGAGTAAATAAGCAGCTGCCA 99164  
QY 603 GCCAGAGTGGCAACCACTCGGATCCCTTCCACTGTGGAAGTTGTTCTTTGCT 662  
DB 99163 GCCAGAGTGGCAACCTGTGAGTTCCTTTCACAGCTGTGTGATGTTGTTCTTCACT 99104  
QY 663 CTTCAATTAATCTTGTGCTGCTCATCTTTGTGTCACACTACTTATGAGCTGA 722  
DB 99103 CTTTGCATTAATCTTGTGCTGCTCATCTTTGTGTCACACTGCTTTATGAGCTGA 99044  
QY 723 ACACCTACCTGCGAGGCTGTGGCTTCATTCCTGAAGTCAC-AGAACAAGAACCACTG 781  
DB 99043 ACACCTACCAATGAAGTCTGAGCTTCACCTCGAAGCTAAGCAAAACGCAACCGCG 98984  
QY 782 GAAGGAACAAGAACTCCCGATGTGCTGCTTAAGAGTGTAAATCACTACGCAAGT 841  
DB 98983 GAAGGAACAAGAACTCCCGATGTG- TGACCTTAAGAGTGTAAATCACTACGCAAGT 98925  
QY 842 CTGACGTTCACTCCGTAAGTCAGTAGACCAACAACCAAGGAAGAACTCTGG 901  
|||||

Db 98924 TTGCAGCTTCATTCTCTG-AGCCAGCGAGACCGGTGAACCCACGAGAAAGAAACTCCGA 98866  
QY 902 AACACCTGAATATTCGAGGAACAATCCAGACACACATCTTTAGAGCTGTAAAC 961  
Db 98865 AACACCTGAATATTCGAGGAACAATCCAGACACACATCTTTAGAGCTGTAAAC 98806  
QY 962 TCACGCGAAGGCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGAAG 1021  
Db 98805 TCACGCGAAGGCTGTGGCTTCATTCTTGAAGTCAGCAAGACCAAGAACCCACCGAAG 98751  
QY 1022 GAACAATTCAGACACAGTAGAGAACTGTATTTTTCATCTGTGGCTTCAGGCTTACT 1081  
Db 98750 -----AACTCAGGACACAGTGTA----- 98732  
QY 1082 CCAGTCATTGAAGTCTCCATTCAGTCCATTAAGGAAACAGAAATGTTTGGAGAGACACA 1141  
Db 98731 -----TTTTCACAGCCGGAATTTGACTTGGAAATCTTAA 98698  
QY 1142 TGTGGAAATGTTATGAGACAGGCTTGAGATGCAATAGGGCATTTCTGATCAAACTAG 1201  
Db 98697 AGTTACTTTTATTTTATGAGAGCTTTGATATCCATTAATAATCTTATACCTCTTTAA 98638  
QY 1202 CTGGAAGCAGGCGCAGGAAATATATCTAAAGAAACAGTTTGTAGACATGATATGC 1261  
Db 98637 TGTATTTTAAATTTTGAATATAGTTTCACTAACAGAAACAAAGATGCACAA----- 98584  
QY 1262 TTTGCATTCGAGCATGTAGATTAATCAAGCAATTAATTAAGAAATATAGCAGGTGCG 1321  
Db 98583 -CAGTACCTTAAACGATTTTCTTATGATCTTAAAGATAGCTATTTTCAGCCAGGCTAG 98525  
QY 1322 ATGGCTCATGCTGTATATCCAGACATTTGGAGGCGGCGGTGTGATCAGAGGTCA 1381  
Db 98524 GTGGCTCAGCCCTGTATATCCAGACATTTGGAGGCGGTGTGATCAGAGGTCA 98465  
QY 1382 GGGCTTGAGACCAAGCTGCGCAACATGTGAAACCCCGTCTCTACTAAAAATACA---- 1437  
Db 98464 GGAGATTGAGACCATCTGTCTAAACGATGAAACCCCGCTCTAGTAAAAATACAAAA 98405  
QY 1438 -----AAATTTAGCTGTGTGTGTGACGACATCTGTAAATCCAG-TACTCAGAGAG 1487  
Db 98404 AAAAAAAAAAAATTTAGCCCGAGTGGGAGGCGGCTGTAGTCTGTGCTACTCGGAG 98345  
QY 1488 GCTGAGCGAGGGAATCTTGTGAATCTTGGAGGCGAGAGTTGACGTAGGCGCAAGATCACA 1547  
Db 98344 GCTGAGCGAGGGAATATGATGAACCCAGAGGCGGAGCTTGCATATAGCGGAGATTGTG 98285  
QY 1548 CCAACAGCATCTCTCTGGGTGACAGGCGAGACTGTGTCTCAAAAAAAAAAAAAA 1604  
Db 98284 CCATTGCACTCCAGCCCTGGGTGACAGTGTGAGACTCATCTCAAAAAAAAAAGATA 98228

XX AEB85185;  
AC 06-OCT-2005 (first entry)  
XX Human phosphodiesterase 4D gene SEQ ID NO:1.  
XX diagnosis; cerebrovascular ischemia; phosphodiesterase 4D; chromosome-5;  
XX cerebroprotective; vasotropic; gene; ds; single nucleotide polymorphism;  
XX SNP.  
OS Homo sapiens.  
FH Key  
FT variation  
FT 78552  
FT location/Qualifiers  
FT /tag= a  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 80127  
FT /tag= b  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 82244  
FT /tag= c  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 83119  
FT /tag= d  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 88456  
FT /tag= e  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 88614  
FT /tag= f  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 90541  
FT /tag= g  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 91022  
FT /tag= h  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 92148  
FT /tag= i  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 93035  
FT /tag= k  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 100795  
FT /tag= l  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 102977  
FT /tag= m  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 104552  
FT /tag= n  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"  
FT 105225  
FT /tag= o  
FT /standard\_name= "single nucleotide polymorphism (SNP)"  
FT /note= "no variation given - see Table 2C"

```
FT variation
FT /tag= p
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 107317
FT /tag= q
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 107849
FT /tag= r
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 108127
FT /tag= s
FT /number= 1 - 4D7-A
FT /note= "see page 47 Table 10A"
FT 108218
FT /tag= t
FT 109301
FT /tag= u
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 111252
FT /tag= v
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 111781
FT /tag= w
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 118914
FT /tag= x
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 120628
FT /tag= y
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "given as N in the sequence"
FT 123312
FT /tag= z
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 123426
FT /tag= aa
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 125304
FT /tag= ab
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 129360
FT /tag= ac
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 129361
FT /tag= ad
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "given as N in the sequence"
FT 130391
FT /tag= ae
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 131865
FT /tag= af
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "given as N in the sequence"
FT 132562
FT /tag= ag
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 133371
FT /tag= ah
FT /standard_name= "single nucleotide polymorphism (SNP)"
```

```
FT variation
FT /note= "no variation given - see Table 2C"
FT 135112
FT /tag= ai
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 135152
FT /tag= aj
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 135641
FT /tag= ak
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 136265
FT /tag= al
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 138456
FT /tag= am
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 142207
FT /tag= an
FT /number= 2 - PDE4D7-1
FT /note= "see page 47 Table 10A"
FT 142329
FT /tag= ao
FT 142556
FT /tag= ap
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
FT 142757
FT /tag= aq
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "no variation given - see Table 2C"
```

Query Match 21.8%; Score 435.6; DB 14; Length 110000;  
Best Local Similarity 67.1%; Pred. No. 3.8e-94;  
Matches 803; Conservative 0; Mismatches 290; Indels 104; Gaps 9;

```
QY 431 TCAGTGGGAGCTTGAGAACTTTGTCTAGCTAAAGATTGTAATGCACCAATCAG 490
DB 99343 TCTGTGGGAGGTAGAACTTTGTCTAGCTGGGGATTGTAACGACCAATCAG 99284
QY 491 CACTGTG-----GTCTAGCTAAAGATTGTAATGCACCAATCAGCTCTGTA 542
DB 99283 CACCTGTCAAAACAGACCAATGACTCTCTGTAACAGCAATCGCTCTTTAA 99224
QY 543 ATGACCAATCAGCAGATGTGGCGGGTCAAAATPAGGAGTAATAAATGCGCACCGA 602
DB 99223 ATGACCAATCAGCAGAGCTGGGTGGGCCAGATAAGAAATPAAAGAGCTGCCAA 99164
QY 603 GCCAGAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTTTGC 662
DB 99163 GCCAGAGTGGGAACCTGCTGAGGTTCCCTTTCATGCTGTGAGCTTTGTTTCA 99104
QY 663 CTTCAAAATTAATCTTGTGCTGCTCATCTTTGTTTCCACACTACTTATAGAGTGA 722
DB 99103 CTTTGAATTAATCTTGTGCTGCTCATCTTTGTTTCCACACTACTTATAGAGTGA 99044
QY 723 ACACTCACTGCGAGGGTCTGTGGCTTCATCTCTGAAGTCAAC-AGACCAAGAACCCACT 781
DB 99043 ACACTCACTGAAGGTCTGCACTTCAACCCCTGAGCTAGCAAAACCGGAACCCGCC 98984
QY 782 GAAGGAACAAGAACTCCGATGTGCTGCTTTAAGAGTGTAACTCACTCGGAAGCT 841
DB 98983 GGAGGAACAAGAACTCCGATGTGCTGCTTTAAGAGTGTAACTCACTCGGAAGGT 98925
QY 842 CTGCAAGCTTCACTCCGGAAGTCAAGTGAACCAACCAACCAAGGAAGAACTCTGG 901
DB 98924 TTGCAAGCTTCACTCTTG-AGCCAGCAAGCCGTAACCAACCAAGGAAGAACTCTCA 98866
QY 902 ACACACTGAATATCTGAAGAAACAATCCAGACAACAATCTTTAGAGCTGTAAAC 961
```

```

Db      98865 ACACATCCAAACATCAGAAAGAAACAACCTCCAGACAGCTGCTTTAAGACTGTAACT 98806
QY      962 TCACCGCAAGGCTGTGGCTTCACTTCTTGAAGTCAGCAAGACCAAGAACCCGGAAG 1021
Db      98805 TCACTGAGAGGCTCTGGCTTCACTTCTTGAATCGTAAGACCAAGAACCCGAC----- 98751
QY      1022 GAACAATTCAGACACAGTAAAGAAATCTGTATTTTGTGATCTGTGGCTTCCAGGGTTACT 1081
Db      98750 -----AATCAGAGACACAGTGA----- 98732
QY      1082 CCAGTCATTGAAGTCTCCATTCAGCTTAAAGAAACAGAGAAATGTTTGGAGGACACA 1141
Db      98731 -----TTTTCACAGCCGGAATAATGACTTGGAAATCTTTA 98698
QY      1142 TGGGAATTTGTAATGACACAGGCTTGAGATGCACTAAGGCACTTCTGATCAAACTGAG 1201
Db      98697 AGTTACTTTTCTTTTATAGCAGCTTTTGTGATGACCAATAAAATCTTATACCTTCTTTAA 98638
QY      1202 CTGGAAGCAGGGCCAGAAATATATATCTAAGAAAGACAGTTTGTGACAGTAATGTC 1261
Db      98637 TGTATTATTAATTTTGAATAAGTTCAGTAACAGAAACAAAGATGACACAA----- 98584
QY      1262 TTTGCACTGACAGCATGTAGATTTCAAGCAATTAATAGAAAAAATATAGCCAGGTGG 1321
Db      98583 -CAGTAACCTTTAAACGATTTTCTTAGATCTTAAAGATAAGTATTTTCAGCCAGGCATG 98525
QY      1322 ATGGCTCATGCGCTGTATCCAGACACTTTGGAGGCGCAAGGGGTGTGATCAGAGGTCA 1381
Db      98524 GTGGCTCAGCGCTGTATCCAGACACTTTGGAGGCGTGAAGGCGGAGATCAGAGGTCA 98465
QY      1382 GCGCTTCAGACAGCGCTGCGCAACATGTGAAACCCCGCTCTACTAAAAATACA---- 1437
Db      98464 GGAAGATGAGACCATCTGGCTTAACAGGTAACCCCGCTCTAGTAAATAACAAAA 98405
QY      1438 -----AAATTTAGCTGTGTGTGTGTCGACGACATCTGTATCCAG-TACTCAGAG 1487
Db      98404 AAAAAAAAAAAATTAAGCCCGAGTGGGGCAGCGCTGTAGTCTCGGCTACTCGGGAG 98345
QY      1488 GCTGAGGCGAGGGAATCTCTTGAATCTTGGAGGCGAGGTTGAGTGAAGCAAGATCACA 1547
Db      98344 GCTGAGGCGAGGGAATTAAGATTAACCCAGAGGCGAGGCTTGAATTAAGCCGAGATTGTG 98285
QY      1548 CCACAGCACTCATCTCTGCTGTGACAGACGAGACTCTGTCTCAAAAAAAAAAAAAA 1604
Db      98284 CCATTGCACCTCCAGCTGGGTGACAGTGAAGTCACTCAAAAAAAAAAAAAAGATA 98228

```

## RESULT 27

ADL27128  
ID ADL27128 standard; DNA; 28670 BP.

AC ADL27128;

DT 20-MAY-2004 (first entry)

DE Human genomic sequence for IQGAP1.

KW Human; ds; gene; CA gene; carcinoma associated gene; cytoskeletal; cancer;

KW carcinoma; lymphoma.

OS Homo sapiens.

XX US2003216558-A1.

XX 20-NOV-2003.

XX 20-DEC-2001; 2001US-00034650.

XX 22-DEC-2000; 2000US-00747377.

XX 02-MAR-2001; 2001US-00798586.

XX (MORRIS D W.

PA (ENGELHARD E K.

```

XX      XX Morris DW, Engelhard EK;
PI      PI WPI: 2003-902052/82.
XX      XX
DR      DR
XX      XX New carcinoma associated gene, useful for preparing a composition for
PT      PT diagnosing or treating carcinoma.
XX      XX
PS      PS Claim 1; SEQ ID NO 16; 342pp; English.
XX      XX
CC      CC The invention relates to a new recombinant nucleic acid (from a CA,
CC      CC carcinoma associated, gene) appearing as ADL27113-ADL27112. Also included
CC      CC are a host cell comprising the recombinant nucleic acid or expression
CC      CC vector, an expression vector comprising the recombinant nucleic acid, a
CC      CC recombinant protein (a carcinoma associated protein) comprising the
CC      CC sequence encoded by the nucleic acid, a method for screening drug
CC      CC candidates, a method for screening for a bioactive agent capable of
CC      CC binding to (or modulating the activity of) a carcinoma associated
CC      CC protein, a method for evaluating the effect of a candidate carcinoma
CC      CC drug, a method of diagnosing carcinoma, a method for inhibiting the
CC      CC activity of a carcinoma associated protein, a method of treating
CC      CC carcinomas, a method of neutralising the effect of a carcinoma associated
CC      CC protein, a biochip comprising one or more nucleic acid segments of the
CC      CC nucleic acid, a method of diagnosing carcinoma or propensity to carcinoma
CC      CC and a method of determining carcinoma associated gene copy number. The
CC      CC nucleic acid is useful for preparing a composition for diagnosing or
CC      CC treating carcinoma especially lymphomas. The present sequence is the
CC      CC genomic sequence from a human carcinoma associated gene.
XX      XX
SQ      SQ Sequence 28670 BP; 7471 A; 6745 C; 6201 G; 8252 T; 0 U; 1 Other;
Query Match 21.7%; Score 433.4; DB 11; Length 28670;
Best Local Similarity 88.6%; Pred. No. 8.3e-94;
Matches 504; Conservative 0; Mismatches 61; Indels 4; Gaps 3;
QY      QY 476 AATGACCAATCAGACCTCTGTCTACTTAAAGATTTGAATCAACCAATCAGCACT 535
Db      23250 AAACGACCAATCAGACCTCTGTCTACTTAAAGATTTGAATCAACCAATCAGCACT 23309
QY      QY 536 CTGTAAATATGACCAATCAGACGATGTGGCGGGGTCAATTAAGGAGTAAATATGGC 595
Db      23310 CTGTAAATATGACCAATCAGACGATGTGGCGGGGTCAATTAAGGAGTAAATATGGC 23369
QY      QY 596 CACCCGAGCAGAG-TGGCAACCACTCGGCTCCCTTCACTGTGGAAGCTTTGT 654
Db      23370 CAGCGAAGCAGAGCAAGCAACCACTCGGCTCCCTTCACTGTGGAAGCTTTGT 23429
QY      QY 655 CTTTGTCTTCAATTAATCTTGTCTGTCTCATTTCTTTGTGTCAACATACCTTTAT 714
Db      23430 CTTTGTCTTCAATTAATCTTGTCTGTCTCA--CTTGTGTGTGTCACTTCAAT 23487
QY      QY 715 GAGCTGTAACATCAGTGGAGGGTCTGTGGCTTCAATCTGTAAGTCAACA-GACCAAGA 773
Db      23488 GAGCTGTAACATCAGTGGAGGGTCTGTGGCTTCAATCTGTAAGTCAACAAGACCA 23547
QY      QY 774 ACCCACTGGAAGGAACAAAGAACTCCGATGTGCTCCTTTAAGCTGTAACTCACT 833
Db      23548 ACCCACTGGAAGGAACAAAGAACTCCGATGTGCTCCTTTAAGCTGTAACTCACT 23607
QY      QY 834 GCGAAGCTTGACGCTTCACTCTGTAAGTCAAGTGAACCAAAACCAACCAAGGAAGA 893
Db      23608 GCGAAGGTGTGCGGCTTCACTCTGTAAGTCAAGTGAACCAAAACCAACCAAGGAAGA 23667
QY      QY 894 AACTGGAACACACCTGAATATCTGAAGGAACAACTCCGACACACCAATCTTTAGAGC 953
Db      23668 AACTGGAACACACCTGAATATCTGAAGGAACAACTCCGACACACCAATCTTTAGAGC 23727
QY      QY 954 TGTAACTCAACCGCAAGGCTGTGGCTTCACTTCTTGAAGTCAAGCAAGCAAGAACCC 1013
Db      23728 TGTAACTCAACCGCAAGGCTGTGGCTTCACTTCTTGAAGTCAAGCAAGCAAGAACCC 23787
QY      QY 1014 ACCGAAAGGAACAAATTCAGACACAGTA 1042

```



Db	23788	ACCAAAAGGATTAATTCAGACCCACTA	23816
		RESULT 28	
		ADQ20461/c	
XX	ID	ADQ20461 standard; DNA; 166181 BP.	
XX	AC	ADQ20461;	
XX	DT	26-AUG-2004 (first entry)	
XX	DE	Human soft tissue sarcoma-upregulated DNA - SEQ ID 3281.	
XX	KX	soft tissue sarcoma; cytosstatic; gene therapy; vaccine; screening; human;	
XX	KW	ds.	
OS		Homo sapiens.	
XX	PN	WO2004048938-A2.	
XX	PD	10-JUN-2004.	
XX	PF	26-NOV-2003; 2003WO-US038193.	
XX	PR	26-NOV-2002; 2002US-0429739P.	
XX	PA	(PROT-) PROTEIN DESIGN LABS INC.	
XX	PI	Aziz N, Ginsburg WM, Zlotnik A;	
XX	DR	WPI; 2004-441208/41.	
XX	PT	Early detection of soft tissue sarcoma comprises determining expression	
XX	PT	of a gene in a first soft tissue sample and a normal soft tissue sample	
XX	PT	and comparing the gene expression, also useful in treating soft tissue	
XX	PT	sarcoma.	
XX	PS	Example 2; SEQ ID NO 3281; 210bp; English.	
XX	XX		
CC	CC	The invention relates to a novel method for detecting soft tissue sarcoma	
CC	CC	which comprises obtaining a first soft tissue sample from an individual	
CC	CC	and a normal soft tissue sample from the same or different individual,	
CC	CC	determining the expression of a gene in both samples and comparing the	
CC	CC	expression of the gene in both soft tissue samples, where a higher level	
CC	CC	of protein expression in the first soft tissue sample indicates the	
CC	CC	presence of soft tissue sarcoma. The method of the invention has	
CC	CC	cyrostatic applications and may be useful for detecting soft tissue	
CC	CC	sarcoma, possibly via gene therapy or vaccine production. The nucleic	
CC	CC	acid sequences may be useful in diagnostic and screening applications.	
CC	CC	The current sequence is that of a human soft tissue sarcoma-upregulated	
CC	CC	DNA of the invention. The current sequence is not shown within the	
CC	CC	specification per se but was submitted in CD format by the inventor.	
XX	XX		
SO		Sequence 166181 BP; 37688 A; 45326 C; 45775 G; 37392 T; 0 U; 0 Other;	
		Query Match 21.2%; Score 425.2; DB 12; Length 166181;	
		Best Local Similarity 66.2%; Pred. No. 1,4e-91;	
		Matches 797; Conservative 0; Mismatches 338; Indels 69; Gaps 10;	
OY	431	TCAGGTGGGGAATTGGAACTTTGTGTCTAGCTAAAGGATTGTAATGACCAATCAG	490
Db	91497	TCTAGTGGGGGATGGAGAACCTTTGTGTCAAGCTCAGGATTGTAATGCAACATCAG	91438
OY	491	CACGTGTGTCTACTCTAAAGATTGTAATGCAACCAATCGACTCTGTAAATGACCA	550
Db	91437	CGCCCTTGCA-----AAACGACCAATCGGCTCTGTGTAATGACCA	91394
OY	551	ATCAGCAGGATGTGGGGGGGTCAATTAAGGAGTAATAAAGTGGCCAGCCGACGAG	610
Db	91393	ATCAGCAGGATGTGGGTGGGGCCAGACGAAGAATAAAGCAGGCTGCCCGCAGCTC	91334
OY	611	TGGCAACCACTGGGGTCCCTTCACACTGTGGAACTTTGTTCTTTTCTCTTCAAA	670

Db	91333	TGGCAACCCCGTGGGGTCCCTTCCANCGCTGTGGAAAGCTTGTCTTTGGCTCTTTGGCA	91274
QY	671	TAAATCTTGCTGCTGCTCATTTCTTTGTGTGCACACTACTTTATGAGCTGTAACTCTAC	730
Db	91273	TAAATCTTGCTGCTGCTCATCTTTGGGGTCCACACTGCGCTTTATGAGCTGTAACTCTAC	91214
QY	731	TGGCAAGGCTCTGTGGCTTCAATTCTGTAAATCA-ACGACCAAGAACCCCACTGGAAAGAAC	789
Db	91213	CACGAAAGGCTGTGACGCTTCACTCTGTGAAGCAATCCATATMACCAAGAGAGGAA	91154
QY	790	AAAGACTCCCGAATGTCTGCTCTTTAAGACTGTAAACTCACTGCGAAGCTGTGACGT	849
Db	91153	GAAACAATGGGGAAGCGCTCTCC-TTAAAGACCGTAACTCACTCAACAAAAGCTGTGACGT	91095
QY	850	TCACCTCTGTAAGTCAGTGAGACCAAAACCAACAGAGAGAGAACTCTGTGACACCT	909
Db	91094	TCACCTCTG-AGCCAGGAGAACCGCGAACCCACAGAGAGAGAACTCCAAACATCC	91036
QY	910	GAATATCTGAAAGAACAACTCCAGACACCAATCTTTCAAGAGCTGTAACTACCGCA	969
Db	91035	GAACATCAAGAAAGGAACAACTCCAGACACCGCCCTTTAAGACTGTAA-----TAATCGG	90986
QY	970	AGGGCTGTGGCTCATTTCTGAAAGTACAGAAACCAAGAACCCAC---GGAAGAAACA	1026
Db	90979	AGGGCTGTGGCTCATTTCTGAAAGTACAGAAACCAAGAACCCCAATCCGAGACACA	90920
QY	1027	AATTCAGACACAGTAGAATCTGTAATTTTGTATCTGTGTGCTTCAGGGTATCTCACT	1086
Db	90919	ATACTTTGAACCACTAGACTATCTGTGTGTGTGCAAAATATACCTTAGGAAATTTTGCAT	90866
QY	1087	CATTGAAGTCTCCATTGACAGCTTTAAGAAACAGAAATGTTTGGAGAGACATGTGG	1146
Db	90859	CCGCAAGTGAATTCATGCTGAAACAGTAATTTGTCAAGTGCCTGCAACAGCTCTTAACT	90800
QY	1147	GAATTTGTATGAGACCAAGCTTGAAGTACATAGGGCAATTTCTGATCAAACTAGTGTGA	1206
Db	90799	GGTTTCTCATCTTCCAAACCCCTCCCTGCTCAATTCATCTTTGG-CACTTTTGAATCTTA	90743
QY	1207	AGCAGGGCCAGAAATTAATCTAAGAGAGACAGTTTGTGTGACAGTAGTAGTCTTGTGC	1266
Db	90742	GGGAATCTTCTGAGGACAGAGTTTAATTAATGATCTTCTCTCAAAAACCTTGGGTGG	90683
QY	1267	ATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAAATATAGCCAGTGCATGTGC	1326
Db	90682	CTCCCTACTAATTTAGAGAGATCCAAACTCTT-----GGCTAAGGCCAGGCATATGTGCG	90629
QY	1327	TCATGCTCTGTATCTCCAGACACTTGGGAGGCCAAAGGGGTGTGATCAAGGTCAGGCT	1386
Db	90628	TCACCTCTGTAAATCCC-----CAACCCAGAAAT	90602
QY	1387	TCGAGACACAGCTGTGGCCAAATGATGTGAACCCCGCTCTACTAAAAATACAAAATTTAGC	1446
Db	90601	TTGAGACAAAGCCGCGCCAAATGTGTAAACCTGTCTTACTTAAATAATACAAAATTTAGC	90542
QY	1447	CTGTGTGTGTGACACGATCTGTAAATCCA-GTACTCAGAGGCTGTGAGCAGGGGAATCT	1505
Db	90541	CAGCTGTGTGTGGGGGTGCTGTAAATCCAGAGTACTTAAGAGGCTGTGAGCAGAGATCT	90482
QY	1506	CTTGAACCTTGGAGAGCAGAGGTTTGCAGTGAGCCAGATCAACACACAGACTTCATCTCTG	1565
Db	90481	CTTGAACCCAGAGGCGAGAGGTTTGCAGTGAGCCAAAGGGCAACCACTGACATCCAGTCTG	90422
QY	1566	GGTGACAGAGCGAGACTGTGTCAAAAAAAGAAAAAAGAAAAAGAAATATATAT	1625
Db	90421	GGCAACGTGTGAGACTCTTATCTCAAAATATATATATATATATATATATATATACAG	90362
QY	1626	CAG 1629	
Db	90361	CTAG 90358	

ID ADQ18633 standard; DNA; 166181 BP.  
XX  
AC ADQ18633;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1452.  
XX  
KM soft tissue sarcoma; cyclostatic; gene therapy; vaccine; screening; human;  
XX  
KW ds.  
XX  
OS Homo sapiens.  
XX  
PN MO2004048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnik A;  
XX  
DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 1452; 210pp; English.  
XX  
CC The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual,  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 166181 BP; 37688 A; 45326 C; 45775 G; 37392 T; 0 U; 0 Other;  
XX  
Query Match 21.2%; Score 425.2; DB 12; Length 166181;  
Best Local Similarity 66.2%; Pred. No. 1,4e-91;  
Matches 797; Conservative 0; Mismatches 338; Indels 69; Gap# 10;  
XX  
QY 431 TCAGTGGGAGCTTGAGAACTTTGTCTAGCTTAAGATTGTAATGACCAATCAG 490  
DB 91497 TCTAGTGGGAGATGGAGAACTTTGTCTAGCTCAAGCTCAAGATTTAATGACCAATCAG 91438  
QY 491 CACTTGTGTCTAGCTTAAGATTGTAATGACCAATCAGACTCTGTAATGACCA 550  
DB 91437 CGCCTGTCA-----AAACAGACCACTCGGCTCTGTGTAATGAGACCA 91394  
QY 551 ATCAGCAGGATGTGGGCGGGGTCAAAATGAGGAGTAAAACTGGCCACCGAGCCACAG 610  
DB 91393 ATCAGCAGGATGTGGGCGGGCCAGACAAAGAAATAAAGCAGCTGCCCAAGCACTC 91334  
QY 611 TGGCAACCACTGGGGTCCCTTCCACACTGTGAAAGCTTTGTTTGTCTTCAAA 670  
DB 91333 TGGCAACCCGCTGGGCTCTCTTCCATGCTGTGAAAGCTTTGTTTGTCTTGTGCA 91274  
QY 671 TAAATCTTGTGCTGCTCATTTCTTGTGTCCACTACTTTATGAGCTGTACACTCAC 730  
DB 91273 TAAATCTTGTGCTGCTCATTTCTTGTGTCCACTGTGCTTTATGAGCTGTACACTCAC 91214

QY 731 TCCGAGGGTCTGTGGCTTTCATTTCTGAAATCA-ACAGACCAAGAACCACTGGAAGAAC 789  
DB 91213 CACGAAGGCTGTGACACTTCACTCTGAAGCAATCCAGACCAATGAAACCAAGGGAGAAA 91154  
QY 790 AAAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCACTCGAAGCTCTGACGT 849  
DB 91153 GAAACAACCTGGACCGCTCTCC-TTAAAGCCGTAACTCAACAAGGCTGTGAGCT 91095  
QY 850 TCACCTCGGAATCAGTGAAGACCAACAAACCCACAGAAAGAAACTGTGACACACT 909  
DB 91094 TCACCTCTTG-AACCAAGGACCGGCAACCCACAGAGAAAGAACTCCAAACATC 91036  
QY 910 GAATATCTGAAGAAACAACTCCAGACACACCATTTTCAAGCTGTAACTCACTCCGA 969  
DB 91035 GAAACATGAAAGAAACAACTCCAGACACGCGCTTTAAGAACTGTAA-----TATCGG 90980  
QY 970 AGGCTCTGTGGCTTTCATTTCTTGAAGTCAGCAAGACCAAGAACCCACC-----GGAAGAA 1026  
DB 90979 AGGCTCTGTGGCTTTCATTTCTTGAAGTCAGTGAAGACCAAGAACCAATTCGGAACA 90920  
QY 1027 AATTCAGACACAGTAGAAGAAATCTGTATTTTGTATCTGTGCTTCAGGGTTACTCCAGT 1086  
DB 90919 AATCTTGAACCATGACTATCTCTGTTGCTGACAAATTAATACCTTAGGATTTTGAT 90860  
QY 1087 CATTAAGTCTCCATTTGACGCTTAAGAAACAGAAATGTTTGAGAGACATGTGG 1146  
DB 90859 CCGCAGATTCATGCTGAACCAAGTTATGCTACAGGCTCGAACAGTCTTTAAT 90800  
QY 1147 GAATTTTATGACCAAGGCTTGAAGATGACATAGGGCAATTTCTGATCAAACTGAGTGA 1206  
DB 90799 GGTTCCTACCTTCAACCCCTCCGCTCCCAATTCATCTTGG-----CATCTTGTATCTA 90743  
QY 1207 AGCAGGGCCAGGAATATATATATATAGAGACAGTTTGTGACAGTATGATCTTTCG 1266  
DB 90742 GGGAACTTCTGAGGACAGGTTTATTAATGATGACTTCTCTCAAAAACCTTCGATGG 90683  
QY 1267 ATCTGAGACATGTAGATTATCAAGCAATTAATTAGAAAAATATAGCCAGGTGCAATGGC 1326  
DB 90682 CTCCTACATTTTATGAGAAAGTCCAAACTCTCTT-----GAGTAAGGCCAGGCTAGTGGC 90629  
QY 1327 TCATGCTGTATATCCAGCACTTTGGAGAGCCAAAGGGGTGTGATCAAGATCAGGCT 1386  
DB 90628 TCACCTCTGTAAATCCC-----CAGCCAGGAAT 90602  
QY 1387 TCGAGACCAAGCCGAGCAATGATGAACCCCGCTCTACTTAAATACAAAATTTAGC 1446  
DB 90601 TTAGACAAAGCCCGGCAACATGTGAACCTGTCTCTTAATAAATACAAATTTAGC 90542  
QY 1447 CTGCTGTGTGGACACCATCTGTATATCCA-GTACTCAGAGGCTGAGGCAAGGAAATCT 1505  
DB 90541 CAGGTGTGTGTGGGAGGCTGTATATCCAGGTACTTAAGAGGCTGAGGCAAGAAATCT 90482  
QY 1506 CTGGAACCTGGAGGACAGAGGTTTGCAGTGAAGCCAAATATCACACAGCACTCAATCTTG 1565  
DB 90481 CTGGAACCCAGAGGAGGAGGTTTGCAGTGAAGCCAAAGGCACTGCACTCACTCACTTG 90422  
QY 1566 GGTGACAGAGCCGAGACCTGTCTCAAAAAAAGAAAAAAGAAAGAAATATTAAT 1625  
DB 90421 GGTGACAGTGTGAGCTTTATCTCAAAATATATATATATATATATATATTAACAGG 90362  
QY 1626 CAAG 1629  
DB 90361 CTAG 90358  
XX  
RESULT 30  
ACN44046/c  
ID ACN44046 standard; DNA; 260027 BP.  
XX  
AC ACN44046;  
XX  
DT 18-NOV-2004 (first entry)

XX Human genomic sequence hCG1735292.  
XX  
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX MO2003073826-A2.  
XX  
XX 12-SEP-2003.  
XX  
XX 28-FEB-2003; 2003WO-US006235.  
XX  
XX 01-MAR-2002; 2002US-00087192.  
XX  
XX (SAGR-) SAGRES DISCOVERY.  
XX  
XX Morris DW;  
XX  
XX MPI; 2003-328604/31.  
XX  
XX  
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX comprises a nucleotide sequence.  
XX  
XX Claim 1; SEQ ID NO 298; 0pp; English.  
XX  
XX The present invention relates to novel DNA and protein sequences which  
XX are associated with carcinomas. The sequences are useful for: (i) for  
XX screening drug candidates; (ii) for screening of bioactive agent capable  
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX carcinoma including lymphoma. The present sequence is one such CA coding  
XX sequence. Note: This patent is an equivalent to basic patent  
XX US2002182586A1, for which no sequence data was published  
XX  
XX Sequence 260027 BP; 72144 A; 50369 C; 51877 G; 83517 T; 0 U; 2120 Other;  
XX  
XX Query Match 21.2%; Score 424.6; DB 11; Length 260027;  
XX Best Local Similarity 66.0%; Pred. No. 2.3e-91;  
XX Matches 714; Conservative 0; Mismatches 344; Indels 23; Gaps 6;  
XX  
QY 1 AAGGCTTGAATATGAGCAATGAGCTTCAAACTCTTATACCACTCTGAGTTGGGC 60  
DB 177582 AAGGCTTGAATATGAGCAATGAGCTTCAAACTCTTATACCACTCTGAGTTGGGC 177523  
QY 61 GACATGAGCTTCTCCCTTTCTAGATCTGTGAGAGCCATCTTATAGTCGATTTG 120  
DB 177522 AAGATGAGCTTCTCCCTTTCTAGATCTGTGAGAGCCATCTTATAGTCGATTTG 177463  
QY 121 GCCCTGTATTTTAACTCTTGGTCAATTTGTTTCTCTAGAGTGAAGCCATCAAGCT 180  
DB 177462 GCCCTGTATTTTAACTCTTGGTCAATTTGTTTCTCTAGAGTGAAGCCATCAAGCT 177403  
QY 181 ACAGATGATCTTCAATATGATGAGCCCAATGAGCTCAACAACTCTGCGAGGACC 240  
DB 177402 ACAGATGATCTTCAATATGATGAGCCCAATGAGCTCAACAACTCTGCGAGGACC 177343  
QY 241 CCTGAGCGAGCCGCGCTTTCATGAGCTTAAAGAGCTCCCTCTGAGAGCACTAC 300  
DB 177342 CCTGAGCGAGCCGCGCTTTCATGAGCTTAAAGAGCTCCCTCTGAGAGCACTAC 177283  
QY 301 CACTGAGAGGCGCTTCTTCAACCCCTATCCAGAGAGAGTGAAGGAGGTATG-CC 359  
DB 177282 CACTGAGAGGCGCTTCTTCAACCCCTATCCAGAGAGAGTGAAGGAGGTATG-CC 177223  
QY 360 AATCCCAACAGAGCTGGGGGTCTCTGTTTGGAGGGGGATTGAGAGTGAAGCCAGCT 419  
DB 177223 AATCCCAACAGAGCTGGGGGTCTCTGTTTGGAGGGGGATTGAGAGTGAAGCCAGCT 177163

QY 420 GGCGTT-CTGGGTCAGGTGGGGGACTTGAGAACTTTTGTGTAGCTTAAGATTGTAA 478  
DB 177162 GGATTTCTGGGGTGGAGGAGCTTGAGAACTTTTGTGTAGCTTAAGATTGTAA 177103  
QY 479 TGCACCAATCAGCACTCTGTGTCTAAAGATTGAATGACCAATCAGCACTCTG 538  
DB 177102 CACACCAATCAGCACTCTGTGTCTAAAGATTGAATGACCAATCAGCACTCTG 177043  
QY 539 TAAATGAGCAATCAGCAAGATGTGGGGGCTCAATTAAGGAGTAAACTGGCCAC 598  
DB 177042 TAAATGAGCAATCAGCAAGATGTGGGGGCTCAATTAAGGAGTAAACTGGCCAC 176983  
QY 599 CCGAGCGAGAGTGAAGCAACCACTGGGTCCTTCCACTGTGAGAGTTGTTCTT 658  
DB 176982 ACTCGGTAAAGAG--CACCAATCAGCGCTGTGTCTAGCTTAAGATTGTAAATGCAC 176925  
QY 659 TGCTCTTCAATTAATCTTGTGTCTGCTCATTCTTGTGTCCACACTTATGAGC 718  
DB 176924 TATCAGCACTCTGTAAATATGACCAAGTCAG-CACTGTGTCTAGCTTAAGATTGTAA 176866  
QY 719 TGTACACTCAGTCCGAGAGGTCTGTGCTTCAATCTGAGTGAAGTGAAGCAACCA 778  
DB 176865 TGTACACTCAGTCCGAGAGGTCTGTGCTTCAATCTGAGTGAAGTGAAGCAACCA 176806  
QY 779 CTGGAAGAGCAAAAGACTCCGATGTGCTCTTAAAGCTGTAACTACTCTGCGAA 838  
DB 176805 CAGGAGATGAGCGGAGG---GCCAAATTAAGGATTAAGGATTAAGGATTAAGG 176749  
QY 839 GCTCTGAGCTTCACTCTGAGTGAAGTGAAGCAACCAACCCAGAGAGAGAAACTC 898  
DB 176748 GCGAATGAGGTCCTCTTCCAGCTGTGAAGTGTGTTGCTCTTAAATATCTT 176689  
QY 899 TGGACACACCTGATATCTGAGAGCAAACTCCAGACACACATCTTCAAGCTGTAA 958  
DB 176688 GCGCTGTTT-----ACTCTTGTGGTCCGACCACTTAAAGCTGTAA 176644  
QY 959 CACTCAGCGAAGGTCGTGTGCTTCAATCTTGAAGTGAAGCAAGCAAGCAACCGG 1018  
DB 176643 CACTCAGCAAGAGGTCGTGTGCTTCAATCTTGAAGTGAAGCAAGCAAGCAACCGG 176584  
QY 1019 AAGGAAACAATTCAGACAGATGAGAAATCTGATTTTGAATCTGTGCTTCAAGGTT 1078  
DB 176583 AAGGAAACAATTCAGACAGATGAGAAATCTGATTTTGAATCTGTGCTTCAAGGTT 176524  
QY 1079 A 1079  
DB 176523 A 176523  
RESULT 31  
ADL13904  
ID ADL13904 standard; DNA; 164772 BP.  
XX  
XX ADL13904;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX Osteoarthritis-associated polymorphic nucleotide #436.  
XX  
XX db; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;  
XX joint space narrowing; osteophyte development; joint pain;  
XX osteoarthritis; SNP; single nucleotide polymorphism.  
XX  
XX Homo sapiens.  
XX  
XX MO2003054166-A2.  
XX  
XX 03-JUL-2003.  
XX  
XX 19-DEC-2002; 2002WO-US041225.  
XX

PR 20-DEC-2001; 2001US-0342603P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX Jones KA, Schafer A;  
 PI WPI; 2003-559141/52.  
 DR  
 XX Determining susceptibility of an individual to joint space narrowing,  
 PT osteophyte development and/or joint pain comprises identifying whether  
 PT the individual has at least one polymorphism in a polynucleotide encoding  
 PT a protein.  
 XX  
 XX Disclosure; SEQ ID NO 436; 297pp; English.  
 XX  
 XX The invention relates to a method of determining susceptibility of an  
 CC individual to joint space narrowing and/or osteophyte development and/or  
 CC joint pain comprising identifying whether the individual has at least one  
 CC polymorphism in a polynucleotide encoding at least one of the protein  
 CC listed in the specification. The methods, composition and agent are  
 CC useful for modulating the susceptibility of an individual to joint space  
 CC narrowing and/or osteophyte development and/or joint pain that is  
 CC associated with a disease, preferably osteoarthritis. The cell line and  
 CC the non-human animal are useful for screening for an agent for diagnosing  
 CC an individual having susceptibility to joint space narrowing and/or  
 CC osteophyte development and/or joint pain. This sequence corresponds to  
 CC the polynucleotide encoding a protein listed in the specification. (Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences).  
 CC  
 SQ Sequence 164772 BP; 50645 A; 32137 C; 31960 G; 50022 T; 0 U; 8 Other;  
 Query Match 21.0%; Score 420.4; DB 10; Length 164772;  
 Best Local Similarity 78.7%; Pred. No. 28-90;  
 Matches 573; Conservative 0; Mismatches 81; Indels 74; Gaps 3;  
 QY 387 GTTTGGAGGGGATTGAGAGGTGAAGCCAGCTGGGCTTGGGTCAGTGGGAACTTGG 446  
 DB 12172 GTTTCAGGGGTTCCAGTGAAGGTTGAATCTGTCTGGGCTTGGGTCAGTGGGAACTTGG 12231  
 QY 447 AGAAGTTTGTGTCTAGCTAAGAGATTGAATGACCAATCAGCACTGTGTCTAGCT 506  
 DB 12232 AGAAGTTTGTGTCTAGCTAAGAGATTGAATGACCAATCAGCACTGTGTCTAGCT 12291  
 QY 507 AAGAGATTGAATGACCAATCAGCACTGTGT----- 539  
 DB 12292 AAGAGTTTGAACGACCAATCAGCACTTTGTAACCAATCAACGCTGTGTGT 12351  
 QY 540 -----AAAATGACCAATCAGCAG 558  
 DB 12352 TTAGCTTAAGGTTTGTGAATGACCAATCAGCACTGTGTAACCAATCAGCAG 12411  
 QY 559 GATGTGGCGGGGTCAATAAAGAGAGTAAATCGGCCACCCGAGCCAGAGTGGCAACC 618  
 DB 12412 GAAGTGGCGAGGGCCATTAAGGGAATTAAGCTGGCCACTGAGCCAGAGCGGCAACC 12471  
 QY 619 CACTCGGGTCCCTTCCACTGTGGAAGCTTTGTTCTTTGCTTTCACAAATTAATCTT 678  
 DB 12472 TGCTTAGTCCCTTCCA-----TGAAGCTTGTGTTCTTTCGCTTGAACAAATTAATCTT 12526  
 QY 679 GCTGCTGCTCATCTTTGTGTGCACTACTCTTATAGAGCTGTAACACTCAGCGAGGG 738  
 DB 12527 GCTGCTGCTCATCTTTGTGTGCACTACTCTTATAGAGCTGTAACACTCAGCGAGGTG 12586  
 QY 739 TCTGTGCTTCATCTCTGAAGTCA-ACAGACCAAGAACCCCACTGGAAGAAACAAGAACT 797  
 DB 12587 TCTGTGCTTCATCTCTGAAGTCAAGTGAAGAACCAAGAACCCCACTGGAAGAAACAAGAACT 12646  
 QY 798 CCCGATGTGCTGCTTTAAGAGCTGTAACACTCAGTGGAAAGCTCTGCAAGCTTACTCTCT 857  
 DB 12647 CTGGAGCTACCACTTTAAGAGCTGTAACACTCAGTGGAGAGTCTGTGCTTCACTCTCT 12706

QY 858 GAAGTCACTGAGACCAACCAACCCAGAGAGAGAACTCTGGAACACACTGAAATATCT 917  
 DB 12707 TAAAGTCAGCAGACCAACCAACCCAGAGAGAGAACTCTGGAACACACTGAAATATCT 12766  
 QY 918 GAAGGAACAACCTCCAGACACCACTCTTTCAGAGCTGTAACACTCAGCGAAGGATCTG 977  
 DB 12767 GAAGGAACAACCTCCAGACACCACTCTTTCAGAGCTGTAACACTCAGCGAAGGATCTG 12826  
 QY 978 TGGCTTCATCTTGAAGTCAAGCAGCAAGAACCCACCGAAGGAAACAAATTCAGACA 1037  
 DB 12827 CAGCTTCATCTTGAAGTCAAGCAGCAAGAACCCACCGAAGGAAATTAATTCGACA 12886  
 QY 1038 CAGTAGGA 1045  
 DB 12887 CACCAAGGA 12894  
 RESULT 32  
 ID AA229173 standard; DNA; 2834 BP.  
 XX AA229173;  
 AC  
 DT 21-FEB-2000 (first entry)  
 XX  
 XX Targeting sequence-1 homologous to part of 5'non-coding region of G-CSF.  
 XX  
 XX Granulocyte colony stimulating factor; G-CSF; genomic sequence; upstream;  
 KW transcription start site; 5'non-coding sequence; DNA construct;  
 KW targeting sequence-1; regulatory region; marker gene; selection;  
 KW homologous recombination; gene therapy; delivery system; CMV promoter;  
 KW hematopoietic progenitor cell; chemotherapy-induced neutropenia;  
 KW bone marrow transplantation; congenital neutropenic disorder;  
 KW chronic idiopathic disorder; plasmid pGG3; neomycin resistance gene; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 PN W09957291-A1.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 05-MAY-1999; 99WO-US009924.  
 XX  
 PR 07-MAY-1998; 98US-0084649P.  
 XX  
 PA (TRAN-) TRANSKARYOTIC THERAPIES INC.  
 XX  
 PT Treco DA, Hartlein MW, Selden RF;  
 XX  
 DR WPI; 2000-072235/06.  
 XX  
 PT Novel genomic sequences used for treating human diseases and disorders.  
 XX  
 PS Disclosure; Fig 6; 58pp; English.  
 XX  
 XX The present DNA sequence is the first targeting sequence that corresponds  
 CC to nucleotides 1470-4723 of the human genomic sequence from a region  
 CC upstream of the transcription start site of granulocyte colony-  
 CC stimulating factor (G-CSF). A DNA construct comprising this targeting  
 CC sequence is cloned into the plasmid pGG3, upstream of a CMV promoter and  
 CC neomycin resistance gene, that functions as the transcriptional  
 CC regulatory sequence and as selectable marker respectively. Homologous  
 CC recombination of this construct into the host cells, can be used to  
 CC modify the expression of G-CSF. These recombinant cells which express G-  
 CC SF are useful for in vitro production of the protein and gene therapy.  
 CC Such cells may also be used in a delivery system for stimulating the  
 CC proliferation and differentiation of haematopoietic progenitor cells, or  
 CC for other conditions that can be treated with G-CSF, like chemotherapy-  
 CC induced neutropenia, to treat patients undergoing bone marrow  
 CC transplantation, chronic idiopathic and congenital neutropenic disorders  
 CC  
 XX Sequence 2834 BP; 670 A; 759 C; 817 G; 588 T; 0 U; 0 Other;

Query Match 20.6%; Score 412.2; DB 3; Length 2834;  
 Best Local Similarity 65.9%; Pred. No. 5e-89;  
 Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

431 TCAGGTGGGACCTTGGAGAACTTTGTGTC-----TAGCTAAAGATTGTAATGC 481  
 1370 TCTGGTGGGGCTTGGAGAAATTTGTGTGACACTCTGTATCTAGTTAATCTAGTGGG 1429  
 482 ACCAATCAGACCTGTGTGTAGCTAAAGATTGTAATGCAACATCAGCACTCTGTAA 541  
 1430 ACTGGAGAACTTTGTGTGTAGCTCAAGGATTGTAAACGACCAATCAGCGCCGTGTCA 1489  
 542 AA-----TGACCAATCAGCAGATGTGTGGCGGGGTCAAAATGAGAGTAA 587  
 1490 AAACAGACCACTGGGCTTACCAATCAGCAGATGTGTGGGGCCAGATTAAGAAATTA 1549  
 588 AAACGTGGCCACCCGAGCCAGCAGTGGCAACCACTCGGGTCCCTTCACACTGTGGAG 647  
 1550 AAGCAGGCTGCCCGAGCCAGCAGTGGCAACCGCACAGGTCCCTATCCAAATATGSCAG 1609  
 648 CTTTGTCTTTTGTCTCTTCAATTAATCTTGTCTGTCTCATTTCTTGTGTCCACATTA 707  
 1610 CTTTGTCTTTTGTCTGT 1669  
 708 CTTTATGAGCTGTGTAACTCACTCACTGAGGGTCTGTGTGTCTTCAATCTGTAAGTCAAC-AG 766  
 1670 CTTTATGAGCTGTGTAACTCACTCACTGAGGGTCTGTGTGTCTTCAATCTGTAAGTCAAC-AG 1729  
 767 ACCAGCAACCTCACTGTGAGGAAACAAATCTCCGATGTGTGTCTTCAATGAGCTGTAAAC 826  
 1730 ACCAGCAGCCACCGGAGGAAATGAACAACTCCGGCGGGCTCC-TTAAGAGCTTAAAC 1788  
 827 ACTCACTGGGAAGCTCTGCAAGCTTCACTCTGTAAGTCACTGAGAACCAACCAACCCACAGA 886  
 1789 ACTCACTGGGAAGCTCTGCAAGCTTCACTCTGTAAGTCACTGAGAACCAACCCACAGA 1847  
 887 AGGAAAGAACTCTGAGACACCTGTAATCTGTAAGGAAACAAATCTCCAGACACCACTCTT 946  
 1848 AGGAAAGAACTCTGAGACACCTGTAATCTGTAAGGAAACAAATCTCCAGATGACCA-CTT 1906  
 947 TCAGAGCTGTAACTCACTCACTGAGGGTCTGTGTGTCTTCAATCTTGTGAGTCAAGACAA 1006  
 1907 TAAAGACTGTAACTCACTCACTGAGGGTCCGGGCTTCTTGTGAAGTCAAGTGAACCA 1966  
 1007 AGAACCACCGGAGGAAACAAATCTCAAGACACAGTGTGAATCTGTATTTTGTGATCTG 1066  
 1967 AACACTCACCAAGTGTGAGACACAGCCAGAGTTTGAATGAGCTGTGGGCAACATGATG 2026  
 1067 GCTTCAGGGTACTCTCAGTCAATGTAAGTCTCCATTGCAAGCTTGAAGAAACAGAGATG 1126  
 2027 AATATGCTCTCTCTGCAAAAAAATTAACAAAATTTGGCGGACATGTGTCTCGT 2086  
 1127 GTTTGAGAGGACATGTGAGAAATTTGTATGACCAAGCTTGTGATGACATAGGCAAT 1186  
 2087 GCTGTGTGTCCAGCTACCGGGAGGCTTAAAGTGGAGATGCTTGAAGCTT-GGAGAGT 2145  
 1187 TCTGATCAAACTAGCTGAGAGCGGCAAGAAATATTAATCTTAAGAAAGACAGTTTGTG 1246  
 2146 GAAGCTGAGTGTGATGTGTATGTATGTACCAAGCCCTCTAGGCTGTGGGGGCAAGACT---- 2200  
 1247 TAGACAGT 1306  
 2201 ----GAGACCTGT 2256  
 1307 AATATGAGGAGT 1366  
 2257 AATATGAGGAGT 2316  
 1367 TGTGATCAG--GAGGTGAGCGTGTGAGACCAAGCTGTGGCAACATGTGTGAAACCCGCTCTC 1424  
 2317 CGGGTCACTTAAGTGTGAGAGTGTGAGACAGGCTGTGGCAACATGTGAAAGAACCCATCTTC 2376  
 1425 TACTAAAAATATCAAAATTAAGCT-----GCTGTGTGTGAGACCATCTGTAAATCCGA 1476

Db 2377 TTCTAAAAATACAAAATTAAGCCGGCTGTGGGGGCACTGTGTGAGCATGCTGTAAATCCCA 2436  
 QY 1477 G-TACTCAGAGAGCTTACAGGAGGAAATCTTTGAACTTTGGAGGACAGAGTTCAGTGA 1535  
 Db 2437 GCTTACTCAGAGAGCTTACAGGAGGAAATCACTTTGAAACCAAGAGGAGGAGTTCAGTGA 2496  
 QY 1536 GCCAAGATCACACACAGCACTCCATCC-----TGGGTGACAGAGCCGAGACTCTG 1585  
 Db 2497 GCCAGATCTGTGCTATTGACCTCCACCACTCCAGCTGTGGCAACAGAGCCAAACTCTG 2556  
 QY 1586 TCTCAAAAAAAAAAAAAAAAAA 1606  
 Db 2557 TCTTAAAAAAAAAAAAAAAAA 2577

RESULT 33  
 AA229169  
 ID AA229169 standard; DNA; 6235 BP.  
 XX  
 AC AA229169;  
 XX  
 DT 21-FEB-2000 (first entry)  
 XX  
 DE Human G-CSF genomic sequence upstream of transcription start site.  
 XX  
 KW Granulocyte colony stimulating factor; G-CSF; genomic sequence;  
 KW transcription start site; 5' non-coding sequence; DNA construct;  
 KW targeting sequence; regulatory region; marker gene; selection;  
 KW homologous recombination; gene therapy; delivery system; upstream;  
 KW hematopoietic progenitor cell; chemotherapy-induced neutropenia;  
 KW bone marrow transplantation; congenital neutropenic disorder;  
 KW chronic idiopathic disorder; de.  
 KM  
 OS Homo sapiens.  
 XX  
 PN MO9957291-A1.  
 XX  
 PD 11-NOV-1999.  
 XX  
 XX 05-MAY-1999; 99WO-US009924.  
 PF  
 XX 07-MAY-1998; 98US-0084649P.  
 PR  
 XX (TRAN-) TRANSKARYOTIC THERAPIES INC.  
 PA  
 PI Treco DA, Heartlein MW, Selden RF;  
 XX  
 DR WPI; 2000-072235/06.  
 XX  
 PT Novel genomic sequences used for treating human diseases and disorders.  
 XX  
 XX Claim 1, Fig 5; 58p; English.  
 XX  
 PS The present DNA sequence is the human genomic sequence from a region  
 CC upstream of the transcription start site of granulocyte colony-  
 CC stimulating factor (G-CSF). This sequence represents nucleotides -6578 to  
 CC -364 relative to the translation start site. G-CSF gene contains five  
 CC exons and four introns. A DNA construct comprising a targeting sequence  
 CC homologous to the 5' non-coding sequence of G-CSF, a transcriptional  
 CC regulatory sequence that differs from the endogenous G-CSF gene, and a  
 CC selectable marker gene for selection of recombinant cells is generated.  
 CC It can be used by homologous recombination to modify the expression of  
 CC endogenous G-CSF. These recombinant cells which express G-CSF are useful  
 CC for in vitro production of the protein and gene therapy. Such cells may  
 CC also be used in a delivery system for stimulating the proliferation and  
 CC differentiation of hematopoietic progenitor cells, or for other  
 CC conditions that can be treated with G-CSF, like chemotherapy-induced  
 CC neutropenia, to treat patients undergoing bone marrow transplantation,  
 CC chronic idiopathic and congenital neutropenic disorders  
 CC  
 SO Sequence 6235 BP; 1586 A; 1610 C; 1708 G; 1331 T; 0 U; 0 Other;

Query Match	20.6%	Score 412.2	DB 3	Length 6235
Best Local Similarity	65.9%	Pred. No. 6.5e-89		
Matches	805	Conservative 0	Mismatches 358	Indels 58
			Gaps 12	
QY	431	TCAGTGGGAGCTTGAGAACTTTTGTGTC-----TAGCTAAAGATTGTAAATGC	481	
DB	3259	TCTGTGGGGCTTGAGAAATGTTTGTGCACTCTGTATCTAGTTATCTAGTGGGG	3318	
QY	482	ACCAATTCAGACTCTGTGTCTAGCTAAAGATTGTAATGACCAATCGACTCTGTAA	541	
DB	3319	ACGTGAGAACTTTGTGTCTAGCTCAGGATTGTAAACGACCAATCGACCCCTGTGA	3378	
QY	542	AA-----TGACCAATCAGAGAGATGGGGGGGTCAATTAAGAGAGTAA	587	
DB	3379	AAACAGACCACTGGGCTTACCAATCAGAGATGGGGGGGTCAATTAAGAGATTA	3438	
QY	588	AAACTGGCCACCCGAGCCAGCAGATGGCAACCACTGGGTCCTTCCACACTGTGAAG	647	
DB	3439	AAGCAGGCTGCCGAGCCAGCAGATGGCAACGAGTCCCTATCCCAATATGGCAG	3498	
QY	648	CTTTGTTCTTTTGTCTTCAATTAATCTTGTCTGTCTCATCTTTTGTGCCAATTA	707	
DB	3499	CTTTGTTCTTTTGTCTTCAATTAATCTTGTCTGTCTCATCTTTTGTGCCAATTA	3558	
QY	708	CCTTTATGAGCTGTACACTCAGCAGGAGTGTGTGCTTCTTCTGAACTCAAC-AG	766	
DB	3559	CTTTTATGAGCTGTACACTCAGCAGGAGTGTGTGCTTCTTCTGAACTCAAC-AG	3618	
QY	767	ACCAGCAACCACTGGAAGGAAACAAAGAACTCCGATGTGTCTCTTTAAGCTGTAA	826	
DB	3619	ACCAGAGGCCACCGGAGGAAATGAACAATCCGGCGGCTCC-TTAAGACTATTAAC	3677	
QY	827	ACTCAGTGGAGCTCTGCAAGCTTCACTCTGAAATGTAAGTGAACCAACCCACCA	886	
DB	3678	ACTCAGGAGAGTGTGAGCTTCACTCT-CAAGCAGAGAGACCAACCAACCCACCA	3736	
QY	887	AGGAAACCACTGGAACCACTGGAATGTGAAGGAAACCACTCCAGACACCACTCT	946	
DB	3737	AGGAAACCACTGGAACCACTGGAATGTGAAGGAAACCACTCCAGATGACCA-CT	3795	
QY	947	TCAGACTGTAACTCAACCGAAGGCTGTGTGCTTCACTTTTGAAGTGAAGACCA	1006	
DB	3796	TAAAGACTGTAACTCACTGCGAGGCTCGCGGCTTCTTTGAAGTCAAGTGAACCA	3855	
QY	1007	AGAACCCACCGGAAGAAATTCAGACACAGTGAATCTGTATTTTGTATCTGTG	1066	
DB	3856	AGACATCAACGAGTTTGGACACAAAGCTTGAATGAGCTTGAAGCTTGAAGCAATG	3915	
QY	1067	GCTTCCAGGTTACTCAGTCAATTTGAAGTCTTCACTTTCAGAGCTTAAAGAAAC	1126	
DB	3916	AAATGCCCTCTGCAAAAAAATTAACAAAAATTTGGCGGAGCATGTGTCTCCT	3975	
QY	1127	GTTTGGAGAGCAGATGTGGAAATTTTATGACCAAGCTTGAAGTCAATGAGGCT	1186	
DB	3976	GCTGTGTGTCCAGCTACCGGAGGCTTAAAGTGGAGATCGTTAGACT-GGAGGT	4034	
QY	1187	TCTGATCAAACTAGCTGGAAGCAGGGCCAGAAAAATTAATTAAGGAAGACAGTT	1246	
DB	4035	GAAAGCTGAGATGCTGTGATTTGACCAAGCCCTTGAAGCTGGGGACAGACT---	4089	
QY	1247	TAGACAGTAGTCTTGTGATCTGAGACATGTGATTTAACAACATTAATTGAAAA	1306	
DB	4090	-----GAAACCTGTTTCCCTCCCAAAAAAATTTGAACAAAGTAAATGAAGTGC	4145	
QY	1307	ATATAGCAGGTGAGTGTCTCATGCTGTAACTCCAGCACTTTGGAGGCCAAGGCT	1366	
DB	4146	ATATAGCAGGTGAGTGTCTCATGCTGTAACTCCAGCACTTTGGAGGCCAAGGCT	4205	
QY	1367	TGATAC--GAGTCAAGGCTTGAAGCAGCTGTGCAACATGTGTAACCCGCTCTC	1424	
DB	4206	CGGCTCACTTAAGTCAAGAGTGTGAACAGCTGTGCAACATGTGTAACCCGCTCTC	4265	
QY	1425	TACTAAAAATCAAAAAATTAGCT-----GGTGTGTGGCAGCATCTGTATCCCA	1476	

DB	4266	TTCTAAAAATCAAAATTAAGCTGTGTGGGGCAGTGTGAGCATGCTGTAAATCCCA	4325	
QY	1477	G-TACTCGAGGGCTGAGGAGGGAATCTCTGAACTTGGAGGACAGAGTTGACGTGA	1535	
DB	4326	GCTACTCGAGGGCTGAGGAGGGAATCACTTGAACCCAGAGGCGGCTTGCAGTGA	4385	
QY	1536	GCCAGATCACACACAGCACTCCATCC-----TGGGTGAAGAGGAGACTCTG	1585	
DB	4386	GCCAGATCTGAGCTTGTGATCCACCACTCCAGCTCCAGGCAACAGGCTAACTCTG	4445	
QY	1586	TCTCAAAAAAAAAAAAAA 1606		
DB	4446	TCTTAAAAAAAAAAAAAA 4466		
RESULT 34				
AA229170	AA229170 standard; DNA; 6679 BP.			
XX	AA229170;			
AC	21-FEB-2000 (first entry)			
XX	Partial sequence of human G-CSF upstream of transcription start site.			
DE	Granulocyte colony stimulating factor; G-CSF; partial genomic sequence;			
KW	transcription start site; 5'non-coding sequence; DNA construct;			
KW	targeting sequence; regulatory region; marker gene; selection;			
KW	homologous recombination; gene therapy; delivery system; upstream;			
KW	haematopoietic progenitor cell; chemotherapy-induced neutropenia;			
KW	bone marrow transplantation; congenital neutropenic disorder;			
KW	chronic idiopathic disorder; ds.			
XX	Homo sapiens.			
OS	Homo sapiens.			
XX	Key			
FH	Location/Qualifiers			
FT	misc_feature			
FT	1..19			
FT	/*tag= a			
FT	/note= "Sequence derived from the junction of genomic			
FT	inset and phage arm in the G-CSF/3 phage clone"			
FT	TATA_signal			
FT	6550..6555			
FT	/*tag= b			
FT	/standard name= "Hogness box"			
FT	6518..6679			
FT	/*tag= c			
FT	/product= "N-terminal end of G-CSF"			
FT	/transl_except= (Pos:6648, AGPATQSPM)			
FT	/note= "no stop codon given"			
XX	MO9957291-A1.			
PN	11-NOV-1999.			
PD	05-MAY-1999; 99MO-US009924.			
PF	07-MAY-1998; 98US-0084649P.			
PR	(TRAN-) TRANSKARYOTIC THERAPIES INC.			
PA	Tresco DA, Heartlein MW, Seiden RF;			
PI	WPI; 2000-072235/06.			
DR	P-PSDB; AAY15190.			
XX	Novel genomic sequences used for treating human diseases and disorders.			
PS	Disclosure; Page 48-49; 58pp; English.			
XX	The present DNA sequence is the human partial genomic sequence from a			
CC	region upstream of the transcription start site of granulocyte colony-			
CC	stimulating factor (G-CSF). A DNA construct comprising a targeting			
CC	sequence homologous to the 5'non-coding sequence of G-CSF, a			



transcriptional regulatory sequence that differs from the endogenous G-CSF gene, and a selectable marker gene for selection of recombinant cells is generated, it can be used by homologous recombination to modify the expression of endogenous G-CSF. These recombinant cells which express G-CSF are useful for in vitro production of the protein and gene therapy. Such cells may also be used in a delivery system for stimulating the proliferation and differentiation of haematopoietic progenitor cells, or for other conditions that can be treated with G-CSF, like chemotherapy-induced neutropenia, to treat patients undergoing bone marrow transplantation, chronic idiopathic and congenital neutropenic disorders

Sequence 6679 BP, 1684 A; 1739 C; 1830 G; 1426 T; 0 U; 0 Other;

Query Match 20.6%; Score 412.2; DB 3; Length 6679;

Best Local Similarity 65.9%; Pred. No. 6,7e-89;

Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

431 TCAGTGGGACTTGGAGAACTTTGTGTC-----TAGCTAAAGATTGTAAATGC 481  
3278 TCTGGTGGGGCTTGGAGAACTTTGTGTCACCTGTATCTAGTTATCTAGTGGGG 3337  
482 ACCAATCAGCCTCTGTCTAGCTAAAGATTGTAAATGCACCAATTCAGCACTCTGTAA 541  
3338 ACGTGGAGAACCTTGTGTCTAGCTCAGGATTGTAAACCAACCAATCAGCGCTGTCA 3397  
542 AA-----TGACCAATCAGCAGATGTGGGGGGGTCAAAATTAAGGAGTAA 587  
3398 AAACAGACCACTGGCTCTACCAATCAGCAGATGTGGGGGGGTCAAAATTAAGGAGTAA 3457  
588 AAACCTGCCACCCGAGCCAGCAGATGTGGCAACCCACTGGGGTCCCTTCCACACTGTGAAG 647  
3458 AAGCAGGCTGCCGAGCCAGCAGATGTGGCAACCCACTGGGGTCCCTTCCACCAATATGCGAG 3517  
648 CTTTGTCTTTTGTCTTCAATTAATCTTGTCTCTCATTTCTTGTCTCAACATTA 707  
3518 CTTTGTCTTTTGTCTTGTGCGATTAATCTTGTCTCTCTGCTTGTGGTCCACATG 3577  
708 CCTTATAGCTGTATACCTCACTGAGGATGTGGTCTTCTTCTGAAGTCAAC 766  
3578 CTTTATAGCTGTATACCTCACTGAGGATGTGGTCTTCTTCTGAAGTCAAC 3637  
767 ACCACGACCACTGTGAAGAACTCCGATGTGCTCTTGAAGCTGTAA 826  
3638 ACCAGAGCCACCCGAGGAAATGAACACTCCGCGCGCTGCC-TTAAAGCTATAAC 3696  
827 ACTCACTGGAGCTCTGCACTTCACTCTCTGAAGTCACTGAGACCAACCAACCAACA 886  
3697 ACTCACTGGAGCTCTGCACTTCACTCTCTGAAGTCACTGAGACCAACCAACCAACA 3755  
887 AGGAAGAACTCTGACACACCTGATATCTGAAGAACTCCAGACACCACTCTT 946  
3756 AGGAAGAACTCTGACACACCTGATATCTGAAGAACTCCAGACACCACTCTT 3814  
947 TCAGAGCTGTAACTCACTCCGCAAGGATGTGGTCTTCTTGAAGTCAAGCAAGCA 1006  
3815 TAAAGACTTAACTCACTCCGCAAGGATGTGGTCTTCTTGAAGTCAAGCAAGCA 3874  
1007 AGAACCACCGAAGAACTTCAGACACAGTGAAGAACTGTATTTTGAATCTGTG 1066  
3875 AGCACTCACAGTTTGGACACAAAGCCGAGATTGAATCAGCTGGGCAACATGATG 3934  
1067 GCTTCCAGGGTAACTCACTCAATTAAGTCTCAATTCAGCTTGAAGAAACAGAGAA 1126  
3935 AAATGCTCTCTGCAAAAAAATAAATAAATAAATTTGCGAGGCTGTGTCTCGT 3994  
1127 GTTGGAGAGACATGTGGAAATTTGTATGACCAAGCTTGAATGATGATGAGGCAAT 1186  
3995 GCTGTGTGTCCAGCTACGCGGAGGCTTAAAGTGGAGATGCTTGAAGCTT-GGAGGT 4053  
1187 TCTGATCAAACTAGCTGGAAGCGGCAAGAAATATATCTTAAGAAACAGATTTTGG 1246  
4054 GAAAGCTGAGTGAAGCTGTGATTTGATCAACAGCCCTTGAAGCTGGGGGACAGACT----- 4108

QY 1247 TAGACAGTAGTAGCTTTTGGACTGTAGACATGTAGATTATCAACATTAATTAGAAAA 1306  
DB 4109 ----GAGACCTGTCTTCCCTCCGCAAAAAAATTGACAAAGTATATAAGAGTGTCTG 4164  
QY 1307 ATATGACCAAGTGTGATGTCTATGCTGTATATCCAGCACTTGGAGGCCAAGGGGTG 1366  
DB 4165 ATATGCTGTAGGCGCAATGTGCTATGCTGTATATCCAGCACTTGGGAAAGCCGAGGG 4224  
QY 1367 TGGATAC--GAGGTAGGCGTTCAGACACAGCTTGGCCCAACATGTGTAAACCCCGTTC 1424  
DB 4225 CGGCTACTTAAGTGAAGAGTGTAGACCAAGCTTGGCCCAACATGTGTAAACCCCGTTC 4284  
QY 1425 TACTAAATAATCAAAATTAAGCT-----GATGTGTGACCACTCTGTATATCCCA 1476  
DB 4285 TTTTAAATAATCAAAATTAAGCTGTGTGGGCAATGTGTGAGCATGCTGTATATCCCA 4344  
QY 1477 G-TACTCAGAGGCTGAGGCAAGGAAATCTTTGAACTTGGAGGAGAGTGTCAATGA 1535  
DB 4345 GCTACTCAGAGGCTGAGGCAAGGAAATCACTTGAACCAAGAGGCGGCTTGCAGTGA 4404  
QY 1536 GCCAAGTCAACACCAAGCACTTCATCC-----TGGTGAACAGCCAGACTCTG 1585  
DB 4405 GCCAAGTGTGCTATGTGATTCACCACTCCAGCTTGGCAACAAAGCCAACTCTG 4464  
QY 1586 TCTCAAAAAAATAAATAAATAA 1606  
DB 4465 TCTTAAAAAATAAATAAATAA 4485

RESULT 35  
AADI17443

ID AADI17443 standard; DNA; 6679 BP.

XX AADI17443;

XX 10-DEC-2001 (first entry)

XX DE

Human granulocyte colony stimulating factor (hugCSF) fragment #1.

XX KW

Human; mutation; homologous recombination; target sequence; gene therapy;

XX KW

homologous recombination-enhancing agent; non-homologous end joining;

XX OS

Homo sapiens.

XX PN

WO200168882-A2.

XX PD

20-SEP-2001.

XX PF

13-MAR-2001; 2001WO-US007870.

XX PR

14-MAR-2000; 2000US-00525160.

XX PA

(TRAN-) TRANSARYOTIC THERAPIES INC.

XX PI

Ivanov E;

XX DR

WPI; 2001-582459/65.

XX PT

Complex or composition comprising a double stranded DNA sequence, a

XX PT

homologous recombination-enhancing agent, and agent inhibiting non-

XX PT

homologous end joining, for promoting alteration of a target sequence in

XX PS

Disclosure; Page 77-79; 82pp; English.

XX CC

The invention relates to a complex for promoting alteration of a target

XX CC

sequence in a cell, comprising a double stranded DNA sequence, a

XX CC

homologous recombination-enhancing agent and an agent inhibiting non-

XX CC

homologous end joining. The invention is used in gene therapy. The

XX CC

complex is useful for promoting an alteration at a selected site of a

XX CC

target sequence of a cell preferably of fungal, plant or animal origin,

or of vertebrate origin which is a primary or secondary mammalian (human)



cell or an immortalised mammalian (human) cell, where target sequence comprises a mutation preferably point mutation having less than 10 base pairs which differ from wild-type sequence, (selected from cystic fibrosis transmembrane regulator (CFTR) gene having mutation changes in an amino acid encoded by codon 508, beta-globin gene having mutation changes in an amino acid encoded by codon 6, Factor VIII gene having mutation changes in an amino acid encoded by codon 2209 or 2229, Factor IX gene, von Willebrand factor gene or xeroderma pigmentosa group G gene); and the DNA sequence comprises a wild-type sequence which can correct the mutation. The method further comprises introducing an agent which inhibits a mismatch-repair protein (expression), which is from Msh2, Msh3, Msh3, Msh1 and PMS2, or is an anti-mismatch-repair protein antibody covalently linked to the DNA sequence, or to Rad52 protein or its fragment. The complex is useful for altering expression of the protein coding sequence of a gene in a cell. The method comprises introducing the complex into the cell, where the DNA sequence comprises a regulatory sequence, maintaining the cell under conditions which permit alteration of a targeted genomic sequence to produce a homologously recombinant cell and maintaining the homologously recombinant cell under conditions which permit expression of the protein coding sequence of the gene under control of the regulatory sequence. Homologously recombinant cell is useful as a vehicle or delivery system for therapeutic proteins, such as enzymes, hormones, cytokines, antigens, antibodies, clotting factors, anti-sense RNA, regulatory proteins, transcription proteins, receptors, structural proteins, novel (non-optimised) proteins and nucleic acid products and engineered DNA and for supplying a therapeutic protein, including erythropoietin, calcitonin, growth hormone, insulin and insulinotropin. The present sequence is human granulocyte colony stimulating factor (hugCSF) fragment used in the invention

Sequence 6679 BP; 1684 A; 1739 C; 1830 G; 1426 T; 0 U; 0 Other;

Query Match 20.6%; Score 412.2; DB 4; Length 6679;

Best Local Similarity 65.9%; Pred. No. 6,7e-89;

Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

431 TCAGTGGGGAGCTTGAGAACTTTGTGTC-----TAGCTAAGATTGTAATGC 481  
 3278 TCTGGTGGGGCTTGGAGAAATGTTGTGACACTGTATCTAGTAATCTAGTGGG 3337  
 482 ACCAATCAGCACTCTGTGTAGCTAAAGATTGTAATSCACCAATCAGCACTGTAA 541  
 3338 ACGTGGAGAACTTGTGTAGCTCAGGATTGTAAACCAATCAGCGCCCTGTCA 3397  
 542 AA-----TGACCAATCAGCAGATGTGGGGGGTCAATAAAGGATAA 587  
 3398 AAACAGACCACTCGGCTCTACCAATCAGCAGATGTGGGGCCGATTAAGAAATAA 3457  
 588 AAATGGCCACCGGAGCAGAGTGGCAACCACTCGGGTCCCTTCCACTGTGAG 647  
 3458 AAGCAGGCTGCCGAGCCAGAGTGGCAACCGCAGAGTCCCTATCCAAATATGGCAG 3517  
 648 CTTTGTCTTTTGGCTCTTCAATTAATCTTGTCTGCTCTATTTTGTGTCACATA 707  
 3518 CTTTGTCTTTTGGCTCTTGTGATTAATCTTGTCTGCTCTTGTGTTGGTCCACACTG 3577  
 708 CTTTATGAGCTGTAACTCACTGCGAGGCTGTGTGCTTCAATCTCTGAAGTAAAC-AG 766  
 3578 CTTTATGAGCTGTAACTCACTGCGAGGCTGTGTGCTTCAATCTCTGAAGTAAAG 3637  
 767 ACCAGGAACCACTCGAAGGAAGAAAGAACTCCGATGTGTGCTTAAAGCTGTAAAC 826  
 3638 ACCAGGAGCCACCGGAGGAGTAAGAACATCCGCGCGCTGCC-TTAAAGCTATTAAC 3696  
 827 ACTCAGTGCAGAGCTTGTGCACTTCACTCTGAAGTCACTGAGAGCAACAAACCAACAGA 886  
 3697 ACTCAGCGGAGGCTGTGCACTTCACTCCT-CAAGCAGGAGACCAAGAACCAACAGAGA 3755  
 887 AGGAAGAACTCTGAGACACCTGTGAATATCTGAAGAAACAAATCTCAGACACAGATTT 946  
 3756 AGGAAGAACTCTGAGACACATCTGAATATCTGAAGAAACAAATCTCAGATCAGCA-CT 3814  
 947 TCAGAGCTGTAACTCAACCCCAAGGCTGTGTGCTTCAATCTTGAATCAGCAAGACCA 1006

3815 TAAAGCTGTAACTCACTCAGAGGTCGCGCTTCTTCTTGAAGTCACTGAGACCA 3874  
 1007 AGAACCCACCGGAAGGAACAAATTCACAGACACAGTGAAGAAATCTGATTTTGAATCTGTG 1066  
 3875 AGCACTTCACTGAGTTTGGACACAGCCAGAGAGTTTGAATCAGCTGGGACACATGATG 3934  
 1067 GCTTCAGAGTTTACTCCAGTCAATGAAGTCTTCATTTGACGCTTAAAGAAACAGAAATG 1126  
 3935 AATCCCTCTCTGCAAAAAAATTAACAAAATTTGGCGGAGCATGTGTCCGT 3994  
 1127 GTTTGAGAGACATGTGGGAATTTTATGACCAAGCTTGAATGACATAGGCAAT 1186  
 3995 GCTTGTGTCCTCAGCTAGCCGAGAGCTTAAAGTGGAGATCTGTAGACCT-GGAGGT 4053  
 1187 TCTGATCAACTAGCTGGAAGCAGAGCCAGGAATTAATTAAGAAAGACAGTTTGTG 1246  
 4054 GAAAGCTGAGTGAAGCTGTGATTTGACACAGCCCTTAGGCTGGGGGACAGACT----- 4108  
 1247 TAGACAGTAGTAGTCTTTGCACTTGAACATGTAGATTATCAAGCAATTAATTAGAAAAA 1306  
 4109 ----GAGACCTGTGTTCCCTCCGCAAAAAATTAACAAAAGTATTAAGAGTGTCTG 4164  
 1307 AATATGCCAGTTCATGCTTATGCTTATGCCAGACTTTGGAGGCCAAGGGGTG 1366  
 4165 AATATGCTAGGCGCAGTGTCTATGCTTATCCAGACTTTGGAGGCCAAGGGGTG 4224  
 1367 TGGATCAC--GAGGTGAGCGTTTCGAGACCACTGGCCCAATGATGTGAAACCCGCTCTC 1424  
 4225 CGGTCACCTTAAGTCAAGAGTGTAGACCAAGCTGGCCCAATGAGAAACCCATCTC 4284  
 1425 TACTAAAAATACAAAAATTAAGCT-----GTTGTGTGACGATCTGTATCCCA 1476  
 4285 TTTTAAAAATACAAATTAAGCTGCTGTGGGGCAGTGTGAGGACTGCTGTATTTCCCA 4344  
 1477 G-TACTCAGAGGCTGAGCAGGGAATCTTTGAATCTTGGAGGAGGTTGACATGA 1535  
 4345 GCTACTCAGAGGCTGAGCAGGGAATCTTTGAATCTTGAATCCAGAGAGGGCGGTTGACATGA 4404  
 1536 GCCAATGACACACAGCACTTCATCC-----TGGGTGACAGAGGAGACTGTG 1585  
 4405 GCCGAGATGTGCTATTCATTCACCACTCCAGCTGGGCAACAGAGCCAAATCTTG 4464  
 1586 TCTCAAAAAAATTAAGGATTA 1606  
 4465 TCTTAAAAAATTAAGGATTA 4485

RESULT 36  
 AAD17447  
 ID AAD17447 standard; DNA; 6753 BP.  
 AAD17447;  
 10-DEC-2001 (first entry)  
 XX  
 DE Human interferon alpha2 (huIFNalpha2) fragment #4.  
 XX  
 KW Human; mutation; homologous recombination; target sequence; gene therapy;  
 KW homologous recombination; enhancing agent; non-homologous end joining;  
 KW therapeutic protein; interferon alpha2; huIFNalpha2; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168882-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PE 13-MAR-2001; 2001WO-US007870.  
 XX  
 FR 14-MAR-2000; 2000US-00525160.  
 XX  
 PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX Ivanov E;  
XX  
XX  
XX WPI; 2001-582459/65.  
XX  
XX  
XX Complex or composition comprising a double stranded DNA sequence, a  
XX homologous recombination-enhancing agent, and agent inhibiting non-  
XX homologous end joining, for promoting alteration of a target sequence in  
XX a cell.  
XX  
XX Disclosure, Page 80-82; 82pp; English.

XX The invention relates to a complex for promoting alteration of a target  
XX sequence in a cell, comprising a double stranded DNA sequence, a  
XX homologous recombination-enhancing agent and an agent inhibiting non-  
XX homologous end joining. The invention is used in gene therapy. The  
XX complex is useful for promoting an alteration at a selected site of a  
XX target sequence of a cell preferably of fungal, plant or animal origin,  
XX or of vertebrate origin which is a primary or secondary mammalian (human)  
XX cell or an immortalised mammalian (human) cell, where target sequence  
XX comprises a mutation preferably point mutation having less than 10 base  
XX pairs which differ from wild-type sequence, (selected from cystic  
XX fibrosis transmembrane regulator (CFTR) gene having mutation changes in  
XX an amino acid encoded by codon 508, beta-globin gene having mutation  
XX changes in an amino acid encoded by codon 6, Factor VIII gene having  
XX mutation changes in an amino acid encoded by codon 2209 or 2229, Factor  
XX IX gene, von Willebrand factor gene or xeroderma pigmentosa group G gene)  
XX; and the DNA sequence comprises a wild-type sequence which can correct  
XX the mutation. The method further comprises introducing an agent which  
XX inhibits a mismatch-repair protein (expression), which is from Msh2,  
XX Msh6, Msh3, Mml1 and PMS2, or is an anti-mismatch-repair protein antibody  
XX covalently linked to the DNA sequence, or to Rad52 protein or its  
XX fragment. The complex is useful for altering expression of a protein  
XX coding sequence of a gene in a cell. The method comprises introducing the  
XX complex into the cell, where the DNA sequence comprises a regulatory  
XX sequence, maintaining the cell under conditions which permit alteration  
XX of a targeted genomic sequence to produce a homologically recombinant  
XX cell and maintaining the homologically recombinant cell under conditions  
XX which permit expression of the protein coding sequence of the gene under  
XX control of the regulatory sequence. Homologically recombinant cell is  
XX useful as a vehicle or delivery system for therapeutic proteins, such as  
XX enzymes, hormones, cytokines, antigens, antibodies, clotting factors,  
XX anti-sense RNA, regulatory proteins, transcription proteins, receptors,  
XX structural proteins, novel (non-optimised) proteins and nucleic acid  
XX products and engineered DNA and for supplying a therapeutic protein,  
XX including erythropoietin, calcitonin, growth hormone, insulin and  
XX insulinotropin. The present sequence is human interferon alpha2  
XX (huIFNalpha2) fragment used in the invention  
XX  
XX

SQ Sequence 6753 BP; 2096 A; 1477 C; 1573 G; 1607 T; 0 U; 0 Other;

Query Match 20.6%; Score 412.2; DB 4; Length 6753;

Best Local Similarity 65.9%; Pred. No. 6.7e-89; Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

QY 431 TGAAGTGGGACCTTGGAGAACTTTGTGTC-----TAGCTAAAGATTGTAAATGC 481  
DB 2779 TCTGGTGGGGCCCTTGGAGAAATGTTGTGTCACACTCTGTATCTATCTAATCTAGTGGG 2838  
QY 482 ACCAATCAGACACTCTGTGTCTAGCTAAAGATTGTAAATGACCAATCAGACTCTGTAA 541  
DB 2839 ACGTGAAGAACTTTGTGTCTAGCTCAGGATTTGAAACGACCAATCAGCCCTGTCA 2898  
QY 542 AA-----TGAACCAATCAGAGATGTGGCGGGGTCAATTAAGGAGTAA 587  
DB 2899 AAACAGACCACTCGGCTTACCAATCAGAGATGTGGGTGGGGCCAGATTAAGAGATAA 2958  
QY 588 AAACAGGACCAACCGAGCGAGTGGCAACCACTCGGGTCCCTTCCACACTGTGGAG 647  
DB 2959 AAGCAGGCTGCGGAGCGAGTGGCAACCGCAAGGTCCTTATCCCAATATATGGCAG 3018  
QY 648 CTTTGTCTTTTGTCTTCAATTAATCTTGTGCTCAATCTTTTGTGTCACACTA 707  
|||||

DB 3019 CTTTGTCTTTTGTCTTGGCATTAATCTGTACTGTGCTCGTTTGGGTCCACACTG 3078  
QY 708 CCTTATGAGCTGTAACTCACTGAGAGGTCTGTGCTTCACTTCTTAAGTCAAC-AG 766  
DB 3079 CTTTATGAGCTGTAACTCACTGAGAGGTCTGTGAGCTTCACTCTTAAGCTCACTAG 3138  
QY 767 ACCAGAACCCACTGGAAGAAACAAAGACTCCCGATGTGCTGCTTTAAGACTGTAA 826  
DB 3139 ACCAGAGCCCAACCGGAGAGATGAACAACTCCGCGCGGTGCC-TTAAAGCTATTAAC 3197  
QY 827 ACTCACTGGAAGCTGTGAGCTTCACTCTGTAAGTCACTGAGACCAAACTCCACAGA 886  
DB 3198 ACTCACCGGAGAGTGTGAGCTTCACTCTCT-CAGCAGAGAGACCAAGCAACCAACAGA 3256  
QY 887 AGGAAGAACTCTGAGACACACTGAAATATCTGAAGAAACAACTCCAGACACACTT 946  
DB 3257 AGGAAGAACTGTGAGACACACTGAAATATCTGAAGAAACAACTCCAGATGACCA-CT 3315  
QY 947 TCAGAGCTGTAACTCACTCACCGCAAGGCTGTGCTTCACTTCTGAAGCAGCAAGCA 1006  
DB 3316 TAAAGCTGTAACTCACTCACTGAGAGGTCCGCGCTTCTTTGAAGTCACTGAGACCA 3375  
QY 1007 AGAACCACCGAAGGAAAGAAATTCAGACACAGTAGAAATCTGATTTTGAATCTGTG 1066  
DB 3376 AGACCTCACCAAGTTTGACACAAAGCCAGAGTTTGAATCAGCTGGGCAACATGATG 3435  
QY 1067 GCTTCCAGGCTTATCTCACTGATCTTGAAGTCTCCATTTGACCTTAAAGAAACAGAGAATG 1126  
DB 3436 AATGCGCTCTCTGCAAAAAAATTTCAAAAATTTGGCGGAGCATGTGTCCTGT 3495  
QY 1127 GTTTGAGAGAGACATGTGGAATTTGTTAGACCAAGCTTGAAGTCAATAGGCAAT 1186  
DB 3496 GCTTGTGTCCAGCTTACCGGAGCTTAAAGTGGAGATGCTTGAAGCTT-GGAGAGT 3554  
QY 1187 TCTGATCAACCTAGCTGTGAGACAGGCGCAGGAAATATATATCTTAAAGAAAGAGTTT 1246  
DB 3555 GAAAGCTGAGAGAGTGTGATTTGACACACAGCCCTTGTAGCTGGGGACAGACT----- 3609  
QY 1247 TAGACATGATGATCTTTGATCTGATGACATGATGATTAATCAAGCAATTAATTAAGAAAA 1306  
DB 3610 ----GAGACCCCTGTTTCCCTCCGCAAAAAAATTTGCAAAAGTATTAAGAGTCTGTG 3665  
QY 1307 AATATGCCAGGAGCATGCTGATGCTGATATCCAGACATTTGGAGAGCCAAAGGGT 1366  
DB 3666 AATATGCTTAGGCGCAGGTGCTATGCTGTATATCCAGACATTTGGAGAGCCGAGGCGG 3725  
QY 1367 TGGATCAC--GAGGTCAAGCTTTCAGACACAGCTGCGCAATGATGTGAACCCCGTCTC 1424  
DB 3726 CGGCTCACTTAAGTCAAGAGTGTGAGACCAAGCTGCGCAATGAGAGAAAGCCCATCTC 3785  
QY 1425 TACTTAAATATCAAAATTTAGCT-----GTTGTGTGCGACGATCTGTATCCCA 1476  
DB 3786 TTTCTAAATATCAAAATTTAGCTGCGGTGTGGGGAGTGTGAGATGCTGTATATCCCA 3845  
QY 1477 G-TACTCGAGAGCTGAGAGGAGGGAATCTTTGAACTTTGGAGCAGAGGTTGACAGTGA 1535  
DB 3846 GCTACTCAGGAGCTGAGAGGAGGGAATCTTTGAACTTTGAACTTTGAGAGGCTTGAAGTGA 3905  
QY 1536 GCCAAGATCACACACAGACATCCATCC-----TGGTGAACAGAGCAGACTCTG 1585  
DB 3906 GCCGAGATGTGCTCATTTGACTTCAACCACTCAGACCTGAGGCAACAGAGCCAACTCTG 3965  
QY 1586 TCTCAAAAAAAAAAAAAAAAAA 1606  
DB 3966 TCTTAAAAAAAAAAAAAAAAA 3986  
|||||

RESULT 37  
ADB62309  
ID ADB62309 standard; cDNA; 2942 BP.  
XX  
XX ADB62309;  
XX

DT 04-DEC-2003 (first entry)  
XX Human cDNA encoding clone FCBBF3000010.  
XX  
XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
XX tissue regeneration; cell regeneration; membrane protein;  
XX signal transduction-related protein; transcription-related protein;  
XX osteoporosis; neurological disease; cancer; tumour.  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 74..643  
XX FT /\*tag= a  
XX FT /product= "Clone FCBBF3000010 protein"  
XX  
XX EP1308459-A2.  
XX  
XX 07-MAY-2003.  
XX  
XX 28-MAR-2002; 2002EP-00007401.  
XX  
XX 05-NOV-2001; 2001JP-00379298.  
XX 25-JAN-2002; 2002US-00350978.  
XX  
XX (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
XX Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y,  
XX  
XX WPI: 2003-450961/43.  
XX P-PSDB; ADB64279.  
XX  
XX New polynucleotides and polypeptides, useful for developing a diagnostic  
XX marker or medicines for regulation of their expression and activity, or  
XX as targets of gene therapy.  
XX  
XX Claim 1; Page; 222pp; English.  
XX  
XX The invention discloses a polynucleotide comprising a sequence selected  
XX from 1970 fully defined nucleotide sequences which encode novel  
XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
XX or its partial peptide, an antibody binding to the polypeptide or peptide  
XX of the polynucleotide, immunologically assaying the polypeptide or peptide  
XX of the polynucleotide by contacting the polypeptide or peptide  
XX with the antibody of the encoded protein, and observing the binding  
XX between the two, a transformant carrying the polynucleotide in an  
XX expressible manner and an antisense polynucleotide. The oligonucleotide  
XX is useful as a primer for synthesising the polynucleotide, or as a probe  
XX for detecting the polynucleotide. The polynucleotides and encoded  
XX proteins are useful as pharmaceutical agents and many disease-related  
XX genes may be included in them, for developing a diagnostic marker or  
XX medicines for regulation of their expression and activity, or as targets  
XX of gene therapy. The genes are involved in tissue and/or cell  
XX regeneration. Membrane proteins, signal transduction-related proteins,  
XX transcription-related proteins, disease-related proteins and genes  
XX encoding them can be used as indicators for diseases (e.g. osteoporosis,  
XX neurological diseases, cancer, tumours. The cDNA may be used to regulate  
XX the activity or expression of the encoded protein to treat diseases. The  
XX sequence presented is a cDNA of the invention. Note: Some of the sequence  
XX data for this patent is not represented in the printed specification, but  
XX is based on sequence information supplied by the European Patent Office.  
XX  
XX Sequence 2942 BP; 785 A; 743 C; 728 G; 686 T; 0 U; 0 Other;  
XX  
XX Query Match 20.0%; Score 401.2; DB 10; Length 2942;  
XX Best Local Similarity 83.7%; Pred. No. 2.3e-86;  
XX Matches 497; Conservative 0; Mismatches 78; Indels 19; Gaps 3;  
XX  
XX 453 TTGTGCTAGCTAAAGATTGTAATGACCAATCAGACTGTGTCTAGCTAAAGGA 512  
XX ||||||||| ||| ||||||||| ||||||||| ||||||||| |||

Db 2116 TCTGTCTAGCTGAAGGTTTGTAAATGACCAATCAGACTCCGTA----- 2162  
Oy 513 TTGTAATGACCAATCAGACTCTGTAAATGACCAATCAGAGATGTGGCGGCT 572  
Db 2163 ----AAAGGAGTGAATGAGTGTGTTTAAATGACCAATCAGAGATGTGGCGGCT 2218  
Oy 573 CAATTAAGGAGTAAATGAGTGTGTTTAAATGACCAATCAGAGATGTGGCGGCT 632  
Db 2219 CAATTAAGGAGTAAATGAGTGTGTTTAAATGACCAATCAGAGATGTGGCGGCT 2277  
Oy 633 TCCACAGTGTGAAGCTTTGTTCTTTGCTTTCAATTAATCTTGTGCTGCTCATTC 692  
Db 2278 TCCACAGTGTGAAGCTTTGTTCTTTGCTTTCAATTAATCTTGTGCTGCTCATTC 2337  
Oy 693 TTGTGTCACACTACTCTTATGAGCTGTAACTCACTGACGAGGCTGTGGCTTCATT 752  
Db 2338 TTGTGTCACACTACTCTTATGAGCTGTAACTCACTGACGAGGCTGTGGCTTCATT 2397  
Oy 753 CCTGAAGTCAAC-AGACACGAAACCACTGAAAGGAAAGAACTCCCGATGTGCTGCC 811  
Db 2398 CCTTAAGTTAAGCAAGACCAAGAAACCAAGGAGGAAAGAACTCCCGGACACACACC 2457  
Oy 812 TTTAAGAGCTGTAACTCACTGCGAAGCTCTGCAAGCTTCACTGCTGAAGTCAAGTAC 871  
Db 2458 TTTAAGAGCTGTAACTCACTGCGAAGCTCTGCTGCTTCACTGCTGAAGTCAAGTAC 2517  
Oy 872 CACAAACCAAGCAAGAAAGAACTCTGGAACAACCTGAATATCTGAAGAAACAAATCTC 931  
Db 2518 CACAAACCAAGCAAGAAAGAACTCTGGAACAACCTGAATATCTGAAGAAACAAATCTC 2577  
Oy 932 CAGACACCAATCTTCAAGCTGTAACTCACTGCAAGGCTGTGCTTCACTTCTTG 991  
Db 2578 TGGACACACCAATCTTCAAGCTGTAACTCACTGCAAGGCTGTGCTTCACTTCTTG 2637  
Oy 992 AAGTCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1045  
Db 2638 AAGTCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2691  
  
RESULT 38  
ID ABZ74190/C  
ID ABZ74190 standard; DNA; 27666 BP.  
AC ABZ74190;  
XX  
XX 12-MAY-2003 (first entry)  
XX  
XX Secreted protein gene 215 genomic fragment HNGM58, SEQ ID NO.1337.  
XX  
XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;  
XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;  
XX acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
XX drug screening; chromosome identification; chromosome mapping;  
XX cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;  
XX antianemic; vulnery; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX W0200277013-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 26-MAR-2002; 2002WO-US009370.  
XX  
XX 27-MAR-2001; 2001US-0278650P.  
XX 12-SEP-2001; 2001US-00950082.  
XX 12-SEP-2001; 2001US-00950083.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM,  
XX  
XX WPI: 2003-040578/03.

XX New human secreted proteins and nucleic acids, useful for detecting or  
PT treating cancer or other hyperproliferative disorders, autoimmune  
PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.  
XX disclosure, Page 1963-1970; 2474p; English.  
XX AB273281-AB273697 represent cDNAs corresponding to 391 human secreted  
CC protein genes, and AB000947-AB001363 represent the proteins they encode.  
CC AB273698-AB274687 represent human secreted protein genomic fragments. The  
CC invention also encompasses antibodies specific for the secreted proteins,  
CC the use of the secreted proteins in drug screening and recombinant  
CC vectors and host cells comprising a nucleic acid of the invention. The  
CC secreted proteins are thought to be involved in biological activities  
CC associated with cellular signalling, cellular differentiation, cell  
CC migration, prohormone activation and neurotransmitter activity. The  
CC secreted proteins, nucleic acids encoding them, antibodies or antibody  
CC fragments specific for the secreted proteins, and modulators of protein  
CC activity are useful for diagnosing or treating cancers or other  
CC hyperproliferative disorders. Additionally, the secreted proteins and  
CC their nucleic acids may also be used in the treatment of autoimmune  
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS  
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote  
CC wound healing. Nucleic acids of the invention may be used for chromosome  
CC identification, chromosome mapping, in gene therapy, for identifying  
CC individuals from minute biological samples, as hybridisation probes, and  
CC as molecular weight markers. The present sequence represents a human  
CC secreted protein genomic fragment referred to in the disclosure of the  
CC invention  
XX  
XX Sequence 27666 BP; 7795 A; 6389 C; 6455 G; 7027 T; 0 U; 0 Other;  
SQ  
Query Match 20.0%; Score 400.8; DB 8; Length 27666;  
Best Local Similarity 80.7%; Pred. No. 6.1e-86;  
Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;  
XX  
XX 461 TAGCTAAGAGATTGTAATGACCAATCAGCACTGTGTCTTAAAGATTGTAAT 520  
DB 22694 TAGCTAAGAGATTGTAATGACCAATCAGCACTGTGTAA-----AAAT 22651  
XX  
XX 521 GCACCAATGACGACTGTGTAATGACCAATCAGCAAGATGCGGGGTCAAAATAG 580  
DB 22650 GCACCAATGACGCTGTGTAATGACCAATCAGCAAGATGCGGGGTCAAAATAG 22591  
XX  
XX 581 GGAATTAAGATGCGGCAACCGGACGCACTGCGGCTCCCTTCCACT 640  
DB 22590 GGAATTAAGATGCGGCAACCGGACGCGGGGCAACCGCTGCGGTAATCTTCAATGCT 22531  
XX  
XX 641 GTGGAAGCTTTGTTCTTTGCTTCAATTAATCTGCTGCTCAATCTTTGTGTC 700  
DB 22530 GTGGAAGCTTTGTTCTTTGCTTGTGCAATTAATCTGCTGCTCAATTTTGTGTC 22471  
XX  
XX 701 CACACTACCTTTTATGAGCTGTAACTCACTGCGAGGGTCTGTGCTTCTGAAGT 760  
DB 22470 TGTACTACCTTTTATGAGCTGTAACTCACTGCGAGGGTCTGTGCTTCTGAAGT 22411  
XX  
XX 761 CAACAGACCAAGAACCCACTGGAAGAACAAAGAACTCCCGATGTGCTGCTTTAAGAC 820  
DB 22410 CAACAGATACCAAGAACCCCGGAGAACAAAGAACTCCCGATGTGCTTAAAGAC 22351  
XX  
XX 821 TGTAACTACCTGAGGAGCTGCGAGCTTCACTCTGGAAGTGAAGAACCAACCC 880  
DB 22350 TGTAACTACCTGCGAAGGCTGCGACTTACTCTGGAAGTGAAGAACCAACCC 22291  
XX  
XX 881 ACCAGAAGAGAAAGAACTGTGACACACTGAATATCTGAAGAACCAACCTCAGACAC 940  
DB 22290 ACCAGAAGAGAAAGAACTGTGACACTGTGAATCATCTGAAGAACCAACCTCAGACAC 22231  
XX  
XX 941 CATCTTTGAGAGTGAAGCACTCAACGCAAGGCTGTGCTTCAATTTTGAAGTACGA 1000  
DB 22230 CATCTTTGAGAACTGCAACCTCACTGCGAGGGTTC-CAAGCTTCAATTTTGAAGTACGCG 22172  
XX  
XX 1001 AGACCAAGAACCAACGGAAGAACAAATTCAGACACAGTAGAATCTGTATTTTGA 1060  
QY

DB 22171 AGACCAAGAACCAACGGAAGAACAAATTCAGACACAGTAGAATCTGTACTGCA 22112  
QY 1061 TCTGTGCTTCCAGGGTT 1078  
DB 22111 ACCTGTGCTTCCAGGGTT 22094  
RESULT 39  
ADA98719/C  
ID ADA98719 standard; DNA; 27666 BP.  
XX  
XX ADA98719;  
AC  
XX  
XX 20-NOV-2003 (first entry)  
DT  
XX  
XX Human secreted protein-related DNA sequence #312.  
DE  
XX human; secreted protein; cardiovascular disorder; arrhythmia;  
KW atherosclerosis; stroke; endocarditis; congestive heart failure;  
KW rheumatic heart disease; cardiomyopathy; hemorroids; varicose veins;  
KW migraine; thrombosis; neural disorder; immune system disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; renal disorder; proliferative disorder; cancer; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO2003004623-A2.  
FN  
XX  
XX 16-JAN-2003.  
PD  
XX  
XX 26-MAR-2002; 2002WC-US009922.  
PF  
XX  
XX 27-MAR-2001; 2001US-0278650P.  
PR  
XX 12-SEP-2001; 2001US-00950082.  
PR 12-SEP-2001; 2001US-00950083.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Rosen CA, Ruben SM;  
PI  
XX  
XX WPI; 2003-247946/24.  
DR  
XX  
XX  
XX  
XX New human secreted polypeptide and nucleic acid molecules, useful for  
PT diagnosing, preventing, prognosticating or treating cardiovascular  
PT disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or  
PT thrombosis).  
XX  
XX  
XX disclosure; SEQ ID NO 828; 1572bp; English.  
PS  
XX  
XX The invention comprises the amino acid and coding sequence of human  
CC secreted proteins. The DNA and protein sequences of the invention are  
CC useful in the treatment of cardiovascular disorders, such as: arrhythmia,  
CC atherosclerosis, stroke, endocarditis, congestive heart failure,  
CC rheumatic heart disease, cardiomyopathy, hemorroids, varicose veins,  
CC migraine, or thrombosis. The DNA and protein sequences may also be used  
CC for treating or preventing: neural disorders, immune system disorders,  
CC muscular disorders, reproductive disorders, gastrointestinal disorders,  
CC pulmonary disorders, renal disorders, proliferative disorders and/or  
CC cancerous diseases. The present DNA sequence is used in the  
CC exemplification of the invention. NOTE: The present sequence is shown on  
CC the WIPO website.  
XX  
XX  
XX Sequence 27666 BP; 7795 A; 6389 C; 6455 G; 7027 T; 0 U; 0 Other;  
SQ  
Query Match 20.0%; Score 400.8; DB 8; Length 27666;  
Best Local Similarity 80.7%; Pred. No. 6.1e-86;  
Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;  
XX  
XX 461 TAGCTAAGAGATTGTAATGACCAATCAGCACTGTGTCTTAAAGATTGTAAT 520  
DB 22694 TAGCTAAGAGATTGTAATGACCAATCAGCACTGTGTAA-----AAAT 22651  
XX

QY 521 GCACCAATCAGCACTCTGTAAATGAGCAATCAGCAGATGTGGCGGGTCAATTAAG 580  
DB 22650 GGACCAATCAGCGCTCTGTAAATGAGCAATCAGCAGATGTGGCGGGTCAATTAAG 22591  
QY 581 GGAATTAATGAGTGGCCAGCCGAGCAGTGGCAATCCATCTGGGTCCCTTCCACACT 640  
DB 22590 GGAATTAATGAGTGGCCAGCCGAGCAGTGGCAATCCATCTGGGTCCCTTCCACACT 22531  
QY 641 GGAAGAGCTTTGTTCTTTGCTCTTCAATTAATCTTGTCTGCTCACTTTCTTGTGTC 700  
DB 22530 GTGAGAGCTTTGTTCTTTGCTCTTCAATTAATCTTGTCTGCTCACTTTCTTGTGTC 22471  
QY 701 CACACTACCTTTATGAGTGTAACTCACTCAGCAGGAGTCTGTGGCTTCACTTCTGAAGT 760  
DB 22470 TGTACTACCTTTATGAGTGTAACTCACTCAGCAGGAGTCTGTGGCTTCACTTCTGAAGT 22411  
QY 761 CAACAGACCAAGAACCTCACTGGAAGAAACAAAGAACTCCCGATGTGCTGCTTTAAGAC 820  
DB 22410 CAGCAATACCAAGAACCTCCCGAGAGAACAACTCCCGATGTGCTGCTTTAAGAC 22351  
QY 821 TGTAACTACCTGAGAGTCTGAGAGTCTCACTCTGGAAGTCACTGAGACCAAAACC 880  
DB 22350 TGTAACTACCTGAGAGTCTGAGAGTCTCACTCTGGAAGTCACTGAGACCAAAACC 22291  
QY 881 ACCAGAGAGAGAACTCTGAGACACTGAAATATCTGAAGAACTCCAGACAC 940  
DB 22290 ACCAGAGAGAGAACTCCAGATCTGTAACATCTGAAAGAAACCGCAGACAC 22231  
QY 941 CATCTTTCAGAGTGTAACTCACTCAGCAGGAGTCTGTGGCTTCACTTCTGAAGTCACT 1000  
DB 22230 CAGCTTTAAGAACTGCAACACTCAGCAGGAGTCTGGAAGTCACTTCTGAAGTCACT 22172  
QY 1001 AGACCAAGAACCCAGCAGGAGAACTCCAGACAGTGAAGAACTGTAATTTTGA 1060  
DB 22171 AGACCAAGAACCCAGCAGGAGAACTCCAGACAGTGAAGAACTGTAATTTTGA 22112  
QY 1061 TCTGTGGCTTCCAGGGTT 1078  
DB 22111 ACCTCTGCTCTCTGGGTT 22094

RESULT 40  
ABZ67756/c  
ID ABZ67756 standard; DNA; 27666 BP.  
XX  
AC ABZ67756;  
XX  
DT 26-MAR-2003 (first entry)  
XX  
DE Human secreted protein encoding genomic DNA SEQ ID NO 1279.  
XX  
KW Human; secreted protein; noctropic; neuroprotective; cyrostatic;  
XX  
KW virulence; dermatological; immunosuppressive; anti-inflamatory; anti-HIV;  
XX  
KW vulnerability; antibacterial; immunomodulatory; antitubercular; antitubercular;  
XX  
KW antitubercular; cancer; antineoplastic; hepatocellular; cerebroprotective;  
XX  
KW antitubercular; antitubercular; antidiabetic; antidiabetic; anticonvulsant;  
XX  
KW antitubercular; antiparasitic; cardiac; immune disorder; infection; vaccine;  
XX  
KW cardiovascular disorder; neurological disease; nephrotoxic;  
XX  
KW gene therapy; gene; ds.  
OS Homo sapiens.  
XX  
XX MO200277186-A2.  
XX  
PD 03-OCT-2002.  
XX  
XX 26-MAR-2002; 2002WO-US009188.  
XX  
PF 27-MAR-2001; 2001US-0278650P.  
XX  
PR 12-SEP-2001; 2001US-00950082.  
XX  
PR 12-SEP-2001; 2001US-00950083.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX  
PI Rosen CA, Ruben SM;  
XX  
XX WPI: 2003-040583/03.  
DR  
XX  
XX  
PT New human secreted proteins encoded by genes contained in cDNA clones  
PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,  
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or  
PT West Nile fever.  
XX  
XX  
PS Disclosure; Page 1946-1953; 2423pp; English.  
XX  
XX  
XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the  
XX encoded secreted proteins (ABP9470-ABP9872) useful for preventing,  
XX treating or ameliorating medical conditions e.g. by protein or gene  
XX therapy. The genes are isolated from a range of human tissues disclosed  
XX in the specification. The nucleic acids, proteins, antibodies and  
XX (ant)agonists are useful in the diagnosis, treatment and prevention of:  
XX (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
XX adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
XX lung or urogenital; (b) immune disorders e.g. Addison's disease,  
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
XX arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
XX myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.  
XX cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
XX bacterial, fungal and parasitic infections  
XX  
SQ Sequence 27666 BP; 7795 A; 6389 C; 6455 G; 7027 T; 0 U; 0 Other;  
XX  
Query Match 20.0%; Score 400.8; DB 10; Length 27666;  
Best Local Similarity 80.7%; Pred. No. 6.1e-86;  
Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;  
XX  
QY 461 TAGCTAAAGATTTGTAATGACCAATCAGCAGTCTGTCTAGTAAAGATTTGAAT 520  
DB 22694 TAGCTAAAGATTTGTAATGACCAATCAGCAGTCTGTCTAGTAAAGATTTGAAT 22651  
QY 521 GCACCAATCAGCACTCTGTAAATGAGCAATCAGCAGATGTGGCGGGTCAATTAAG 580  
DB 22650 GGACCAATCAGCGCTCTGTAAATGAGCAATCAGCAGATGTGGCGGGTCAATTAAG 22591  
QY 581 GGAATTAATGAGTGGCCAGCCGAGCAGTGGCAATCCATCTGGGTCCCTTCCACACT 640  
DB 22590 GGAATTAATGAGTGGCCAGCCGAGCAGTGGCAATCCATCTGGGTCCCTTCCACACT 22531  
QY 641 GGAAGAGCTTTGTTCTTTGCTCTTCAATTAATCTTGTCTGCTCACTTTCTTGTGTC 700  
DB 22530 GTGAGAGCTTTGTTCTTTGCTCTTCAATTAATCTTGTCTGCTCACTTTCTTGTGTC 22471  
QY 701 CACACTACCTTTATGAGTGTAACTCACTCAGCAGGAGTCTGTGGCTTCACTTCTGAAGT 760  
DB 22470 TGTACTACCTTTATGAGTGTAACTCACTCAGCAGGAGTCTGTGGCTTCACTTCTGAAGT 22411  
QY 761 CAACAGACCAAGAACCTCACTGGAAGAAACAAAGAACTCCCGATGTGCTGCTTTAAGAC 820  
DB 22410 CAGCAATACCAAGAACCTCCCGAGAGAACAACTCCCGATGTGCTGCTTTAAGAC 22351  
QY 821 TGTAACTACCTGAGAGTCTGAGAGTCTCACTCTGGAAGTCACTGAGACCAAAACC 880  
DB 22350 TGTAACTACCTGAGAGTCTGAGAGTCTCACTCTGGAAGTCACTGAGACCAAAACC 22291  
QY 881 ACCAGAGAGAGAACTCTGAGACACTGAAATATCTGAAGAACTCCAGACAC 940  
DB 22290 ACCAGAGAGAGAACTCCAGATCTGTAACATCTGAAAGAAACCGCAGACAC 22231  
QY 941 CATCTTTCAGAGTGTAACTCACTCAGCAGGAGTCTGTGGCTTCACTTCTGAAGTCACT 1000  
DB 22230 CAGCTTTAAGAACTGCAACACTCAGCAGGAGTCTGGAAGTCACTTCTGAAGTCACT 22172  
QY 1001 AGACCAAGAACCCAGCAGGAGAACTCCAGACAGTGAAGAACTGTAATTTTGA 1060  
DB 22171 AGACCAAGAACCCAGCAGGAGAACTCCAGACAGTGAAGAACTGTAATTTTGA 22112

QY 1061 TCTGTGCTTCAGGTT 1078  
| | | | |  
Db 22111 ACCTGTGCTTCAGGTT 22094

RESULT 41  
AAL06082  
ID AAL06082 standard; DNA; 32212 BP.  
XX  
AC AAL06082;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 8770.  
XX  
KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241212P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.

17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 05-JAN-2001; 2000US-025678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
XX disclosure; SEQ ID NO 8770; 1297bp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX CC number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX  
SQ Sequence 32212 BP; 8041 A; 7564 C; 7490 G; 9117 T; 0 U; 0 Other;  
Query Match 20.0%; Score 400.8; DB 4; Length 32212;  
Best Local Similarity 80.7%; Pred. No. 6,4e-86;  
Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;  
QY 461 TAGCTAAGAGTGTGTAATGACCAATCAGCACTGTGTGTAGCTAAGATTGTAAT 520  
DB 4974 TAGCTAAGAGTGTGTAAGGACCAATCAGCACTGTGTGTAGCTAAGATTGTAAT 5017  
QY 521 GCACCAATCAGCACTGTGTGTGTAATGACCAATCAGCACTGTGTGTGCTCAATTCCTGTAAGT 580  
DB 5018 GCACCAATCAGCACTGTGTGTGTAATGACCAATCAGCACTGTGTGTGCTCAATTCCTGTAAGT 5077  
QY 581 GAGGTAAACCTGCGCAGCGGAGCAGTGGCAACCCAGTGGGGTCCCTTCCACACT 640  
DB 5078 GAGGTAAACCTGCGCAGCGGAGCAGTGGCAACCCAGTGGGGTCCCTTCCACACT 5137  
QY 641 GTGGAAGCTTGTCTTTCTCTCTCTTCAATAAATCTTGTGCTCATTCCTTGTGTC 700  
DB 5138 GTGGAAGCTTGTCTTTCTCTCTCTTCAATAAATCTTGTGCTCATTCCTTGTGTC 5197  
QY 701 CACACTAATCTTTAGCTGTGTAACCTCACTGCGAGGGTCTGTGGCTTCATTCTGTAAGT 760  
DB 5198 TGTACTAATCTTTAGCTGTGTAACCTCACTGCGAGGGTCTGTGGCTTCATTCTGTAAGT 5257  
QY 761 CAACAGACCAAGAACCCACTGGAAGAAACAAAGAACTCCGAGTGTGCTTTAAAGC 820  
DB 5258 CACCAATATACAGAAACCCCGGAGGAAACAAACCTCCCAACCTTAAAGC 5317  
QY 821 TGTAACTCACTGCGAAGCTCTGACCTTCACTCTGGAAGTCAGTGAACAACCAACCC 880  
DB 5318 TGTAACTCACTGCGAAGCTCTGACCTTCACTCTGGAAGTCAGTGAACAACCAACCC 5377  
QY 881 ACCAAGAGAAACTCTGACACACCTGTAATCTGGAAGAAACAACTTCCAGACAC 940  
DB 5378 ACCAAGAGAAACTCTGACACACCTGTAATCTGGAAGAAACAACTTCCAGACAC 5437

QY 941 CATCTTCAGAGCTGTGTAACCTCACTACCGAAGGCTGTGGCTTCAATCTTGAAGTCAGCA 1000  
DB 5438 CAGCTTTAAGACTCAACACTCACTGCGAGGTC-CCAGCTTCAATCTTGAAGTCAGCG 5496  
QY 1001 AGACCAAGAACCCACCGAAGGAACAATTCCAGACACAGTGAAGTAATGTAATTTTGA 1060  
DB 5497 AGACCAAGAACCCACCGAAGGAACAATTCCAGACACAGTGAAGTAATGTAATTTTGA 5556  
QY 1061 TCTGTGGCTTCCAGGGTT 1078  
DB 5557 ACCTCGCTCCTGGGTT 5574  
RESULT 42  
ABL98647  
ID ABL98647 standard; DNA; 32212 BP.  
AC ABL98647;  
XX  
XX 21-JUN-2002 (first entry)  
XX  
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 3299.  
XX  
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
XX reproductive system disorder; urinary system disorder; gene therapy;  
XX cardiovascular disorder; respiratory disorder; neurological disorder;  
XX gastrointestinal disease; infection; cytostatic; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155317-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001MO-US001329.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205151P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226688P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.



PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.

PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0251989P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 06-DEC-2000; 2000US-0251566P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251907P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX MPI; 2001-483232/52.  
DR  
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful  
PT for preventing, diagnosing and/or treating testicular cancer.  
XX  
XX Disclosure; SEQ ID NO 3299; 766pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 973  
CC human testicular antigens, and fragments of their genomic sequences. The  
CC sequences can be used in the treatment of cardiovascular, urinary system,  
CC reproductive system, immune, respiratory, neurological and  
CC gastrointestinal disorders, infections, and particularly cancer,  
CC especially testicular cancers. The present sequence is a DNA encoding a  
XX protein fragment of the invention  
XX  
SQ Sequence 32212 BP; 8041 A; 7564 C; 7490 G; 9117 T; 0 U; 0 Other;  
  
Query Match 20.0%; Score 400.8; DB 4; Length 32212;  
Best Local Similarity 80.7%; Pred. No. 6.4e-86;  
Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;  
  
QY 461 TAGCTAAAGATTGTAATGACCAATCAGCACTGTGCTTAAAGATTGTAAT 520  
DB 4974 TAGCTAAAGTTTGTAAAGGACCAATCAGCACTGTGTA-----AAAT 5017  
  
QY 521 GCACCAATCAGCACTGTGTAATGACCAATCAGCAATGAGATGTGGCGGCTCAATTAAG 580  
DB 5018 GCACCAATCAGCGCTCTGTAATGACCAATCAGCAATGAGATGTGGCGGCAATTAAG 5077  
  
QY 581 GGAGTAAACCTGGCCAGCCGAGCGAGCGGCAACCCAGTGGGTCCCTCCCACT 640  
DB 5078 GGATTAAGCTGGCCAGCTGCGCAGCGGGGCAACCCGCTGGGTACTCTTACATGCT 5137  
  
QY 641 GTGGAAGTTGTTCTTTTGTCTTTCACATAATCTTGCTGCTCAATCTTTGTGTC 700  
DB 5138 GTGGAAGTTGTTCTTTTGTCTTTCACATAATCTTGCTGCTCAATCTTTGTGTC 5197  
  
QY 701 CACACTACTTTATAGAGCTTAACACTCACTCGAGAGGCTGTGGTTATTCCTGAAT 760  
DB 5198 TGTACTACTTTATAGAGCTTAACACTCACTCGAGAGGCTGTGGTTATTCCTGAAT 5257  
  
QY 761 CAACAGACCAAGCAACCACTGGAAGGAACAAAGAACTCCGATGTGCTCCTTAAGAC 820

Db 5258 CAGCAATACCAAGAACCCCGGAGGAAACAACTCCCACTGTCACCTTTAGAGC 5317  
QY 821 TTTAACTCCTCAGTGGAACTCTGCAAGCTTCTCTGTAAGTCAGAGACCAAAACCC 880  
Db 5318 TTTAACTCCTCAGTGGAACTCTGCAAGCTTCTCTGTAAGTCAGAGACCAAAACCC 5377  
QY 881 ACCAGAGAGAGAACTCTGCAACACCTGATATCTGAAGAAACAACTCCAGACAC 940  
Db 5378 ACCAGAGAGAGAACTCTGCAAGCTTCTCTGTAAGTCAGAGACCAAAACCCAGACAC 5437  
QY 941 CATCTTTCAGAGCTGTAACTCAGCGCAAGGGCTGTGCTTCACTTTGAAGTCAGCA 1000  
Db 5438 CAGCTTTAAGAACTGCAACACTCAGCTGCGAGGGTC-CCAGCTTCATCTTGAAGTCAGCG 5496  
QY 1001 AGACCAAGAACCCACCGGAGAGAAACAAATTCAGACACAGTAAAGATCTGTATTTTGA 1060  
Db 5497 AGACCAAGAACCCACCGGAGAGAAACCAATTCAGACACAGTAAAGATCTGTACTGCA 5556  
QY 1061 TCTGTGCTTCCAGGGTT 1078  
Db 5557 ACCTCTGCTCTCTGGTT 5574

RESULT 43  
ABZ74191/c  
ID ABZ74191 standard; DNA; 41206 BP.  
AC ABZ74191;  
XX 12-MAY-2003 (first entry)  
DE Secreted protein gene 215 genomic fragment HNGM58, SEQ ID NO:1338.  
XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;  
XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;  
XX acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
XX drug screening; chromosome identification; chromosome mapping;  
XX cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;  
XX antianaemic; vulnerary; gene; ds.  
XX Homo sapiens.  
XX WO200277013-A2.  
XX 03-OCT-2002.  
XX 26-MAR-2002; 2002WO-US009370.  
XX 27-MAR-2001; 2001US-0278650P.  
XX 12-SEP-2001; 2001US-00950082.  
XX 12-SEP-2001; 2001US-00950083.  
XX (HDMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM.  
XX WPI; 2003-040578/03.  
XX New human secreted proteins and nucleic acids, useful for detecting or  
XX treating cancer or other hyperproliferative disorder, autoimmune  
XX disorders, inflammatory disorders, HIV disease, hepatitis or anemia.  
XX Disclosure; Page 1970-1980; 2474pp; English.

CC migration, prohormone activation and neurotransmitter activity. The  
CC secreted proteins, nucleic acids encoding them, antibodies or antibody  
CC fragments specific for the secreted proteins, and modulators of protein  
CC activity are useful for diagnosing or treating cancers or other  
CC hyperproliferative disorders. Additionally, the secreted proteins and  
CC their nucleic acids may also be used in the treatment of autoimmune  
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS  
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote  
CC wound healing. Nucleic acids of the invention may be used for chromosome  
CC identification, chromosome mapping, in gene therapy, for identifying  
CC individuals from minute biological samples, as hybridisation probes, and  
CC as molecular weight markers. The present sequence represents a human  
CC secreted protein genomic fragment referred to in the disclosure of the  
XX invention

QY 461 TAGCTAAAGATTTGTAATGACCAATCAGCACTGTGTCTAAGTAAAGATTTGTAAT 520  
Db 36233 TACTTAAAGTTTGTAAAGGACCAATCAGCACTGTGTA-----AAAT 36190  
QY 521 GCACCAATCAGCACTGTGTAATGAGCAATCAGAGATGTGGCGGGTCAATTAAG 580  
Db 36189 GGAACCAATCAGCGCTGTGTAATGAGCAATCAGAGATGTGGCGGGTCAATTAAG 36130  
QY 581 GGAGTAAATCTGGCCACCCGAGCCAGCAAGTGGCAACCCACTGGTCCCTCCACACT 640  
Db 36129 GGATTAATGAGCTGGCCACCTGGCCAGCGGGGCAACCCGCTGGGTACTCTTACAGCT 36070  
QY 641 GTGGAAGCTTTGTTCTTTGCTCTTCAATTAATCTTGCTGCTCAATTTCTTGTGTC 700  
Db 36069 GTGGAAGCTTTGTTCTTTGCTCTTCAATTAATCTTGCTGCTCAATTTCTTGTGTC 36010  
QY 701 CACACTTACTTTATGAGCTGTAACTCACTGAGGAGTCTGTGCTTCAATTTCTTGAAGT 760  
Db 36009 TGTACTTACTTTATGAGCTGTAACTCACTGAGGAGTCTGTGCTTCAATTTCTTGAAGT 35950  
QY 761 CAACAGACCAAGAACCCACTGGAAGAGCAAAAGAACTCCCGATGTCGCTTTAAGAGC 820  
Db 35949 CAGCAATATCCACCAAGAACCCCGGAGAGAAACAACTCCCAAGTGTCCACTTTAAGAGC 35890  
QY 821 TGTAACTCCTCAGTGGAACTCTGCAAGCTTCACTCTGTAAGTCAGAGACCAAAACCC 880  
Db 35889 TGTAACTCCTCAGTGGAACTCTGCAAGCTTCACTCTGTAAGTCAGAGACCAAAACCC 35830  
QY 881 ACCAGAGAGAGAACTCTGCAACACCTGATATCTGAAGAAACAACTCCAGACAC 940  
Db 35829 ACCAGAGAGAGAACTCTGCAAGCTTCACTCTGTAAGTCAGAGACCAAAACCCAGACAC 35770  
QY 941 CATCTTTCAGAGCTGTAACTCAGCGCAAGGGCTGTGCTTCACTTTGAAGTCAGCA 1000  
Db 35769 CAGCTTTAAGAACTGCAACACTCAGCTGCGAGGGTC-CCAGCTTCATCTTGAAGTCAGCG 35711  
QY 1001 AGACCAAGAACCCACCGGAGAGAAACAAATTCAGACACAGTAAAGATCTGTATTTTGA 1060  
Db 35710 AGACCAAGAACCCACCGGAGAGAAACCAATTCAGACACAGTAAAGATCTGTACTGCA 35651  
QY 1061 TCTGTGCTTCCAGGGTT 1078  
Db 35650 ACCTCTGCTCTCTGGTT 35633

RESULT 44  
ADA98720/c  
ID ADA98720 standard; DNA; 41206 BP.  
AC ADA98720;  
XX 20-NOV-2003 (first entry)

XX DE Human secreted protein-related DNA sequence #313.  
XX KW human; secreted protein; cardiovascular disorder; arrhythmia;  
XX KW atherosclerosis; stroke; endocarditis; congestive heart failure;  
XX KW rheumatic heart disease; cardiomyopathy; hemorhoids; varicose veins;  
XX KW migraine; thrombosis; neural disorder; immune system disorder;  
XX KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
XX KW pulmonary disorder; renal disorder; proliferative disorder; cancer; ds.  
OS Homo sapiens.  
PN WO2003004623-A2.  
XX 16-JAN-2003.  
XX 26-MAR-2002; 2002WO-US009922.  
XX 27-MAR-2001; 2001US-0278650P.  
XX 12-SEP-2001; 2001US-00950082.  
XX 12-SEP-2001; 2001US-00950083.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX WPI; 2003-247946/24.  
XX  
XX New human secreted polypeptide and nucleic acid molecules, useful for  
XX diagnosing, preventing, prognosticating or treating cardiovascular  
XX disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or  
XX thrombosis).  
XX  
XX Disclosure; SEQ ID NO 829; 1572bp; English.  
XX  
XX The invention comprises the amino acid and coding sequence of human  
XX secreted proteins. The DNA and protein sequences of the invention are  
XX useful in the treatment of cardiovascular disorders, such as: arrhythmia,  
XX atherosclerosis, stroke, endocarditis, congestive heart failure,  
XX rheumatic heart disease, cardiomyopathy, hemorhoids, varicose veins,  
XX migraine, or thrombosis. The DNA and protein sequences may also be used  
XX for treating or preventing: neural disorders, immune system disorders,  
XX muscular disorders, reproductive disorders, gastrointestinal disorders,  
XX pulmonary disorders, renal disorders, proliferative disorders and/or  
XX cancerous diseases. The present DNA sequence is used in the  
XX exemplification of the invention. NOTE: The present sequence is shown on  
XX the WPIO website.  
XX  
XX Sequence 41206 BP; 11685 A; 9614 C; 9725 G; 10182 T; 0 U; 0 Other;  
XX  
XX Query Match 20.0%; Score 400.8; DB 8; Length 41206;  
XX Best Local Similarity 80.7%; Pred. No. 6,9e-86;  
XX Matches 499; Conservative 0; Mismatches 102; Indels 17; Gaps 2;  
XX  
XX 461 TAGCTAAGAGTATTAATGACCAATCAGCACTCTGTCTAAGATTGTAAT 520  
XX |||||  
XX 36233 TAGCTAAGAGTATTAATGACCAATCAGCACTCTGTAA-----AAAT 36190  
XX  
XX 521 GGACCAATCAGCACTCTGTAAATGACCAATCAGCAAGATGTGGCGGGTCAATTAAG 580  
XX |||||  
XX 36189 GGACCAATCAGCGCTCTGTAAATGACCAATCAGCAAGATGTGGCGGGTCAATTAAG 36130  
XX  
XX 581 GGAGTAAATCTGGCCACCGAGCCAGCAGTGGCAACCACTGGGTCCCTTCCACACT 640  
XX |||||  
XX 36129 GGAGTAAATCTGGCCACCGAGCCAGCGGGGCAACCGCTGGGTACTTTCATGCT 36070  
XX  
XX 641 GTGGAAGCTTTGTTCTTTGCTCTTCAATTAATCTTGCTGTGCTCAATCTTTGTC 700  
XX |||||  
XX 36069 GTGGAAGCTTTGTTCTTTGCTCTTTCAGCAATTAATCTTGCTGTGCTCAATCTTTGTC 36010  
XX  
XX 701 CACACTACTTTATAGCTGTAACTACTGCGAGGGTCTGTGGCTTCAATCTTGAAGT 760  
XX |||||  
XX 36009 TGTACTACTTTATAGCTGTAACTACTGCGAGGGTCTGTGGCTTCAATCTTGAAGT 35950  
XX

QY 761 CAACAGACCAAGAACCACTGGAAGAACAAAGAACTCCGATGTGCTTTAAGAGC 820  
DB 35949 CAGCAATACAGAGAACCCCGGAGAGAACAACTCCCAAGTGTACCTTTAAGAGC 35890  
QY 821 TGTAACTACTCTGCGAAGCTCTGCAAGCTTCACTCTGAAGTCAAGAACCAACCC 880  
DB 35889 TGTAACTACTCTGCGAAGCTCTGCGAAGCTTCACTCTGAAGTCAAGAACCAACCC 35830  
QY 881 ACCAAGAGAGAACTCTGGAACACCTGAAATATCTGAAGAACCAACTCTGAGACAC 940  
DB 35829 ACCAAGAGAGAACTCTGGAACACCTGAAATATCTGAAAGAACCAACCCGAGACAC 35770  
QY 941 CATCTTCAAGAGTGTAACTCAACCGCAAGGCTGTGGCTCATTTCTGAAGTCAAGCA 1000  
DB 35769 CAGCTTTAAGAACTGTCAACCTCACTGCGAGGGTC-CAGACTTCATTTCTGAAGTCAAGCG 35711  
QY 1001 AGACCAAGAACCCAGCGAAGAACAAATTCAGACACAGTAAATCTGTAATTTTGA 1060  
DB 35710 AGACCAAGAACCCAGCGAAGAACCAATTCAGACACAGTAAATCTGTAATTTTGA 35651  
QY 1061 TCTGTGGCTTCCAGGGTT 1078  
DB 35650 ACCTTGCCCTCTGCGGTT 35633  
DB  
RESULT 45  
ABZ67757/C  
ID ABZ67757 standard; DNA; 41206 BP.  
AC ABZ67757;  
XX  
DT 26-MAR-2003 (first entry)  
XX  
XX Human secreted protein encoding genomic DNA SEQ ID NO 1280.  
XX  
XX Human; secreted protein; nootropic; neuroprotective; cytostatic;  
XX virucide; dermatological; immunosuppressive; anti-inflamatory; anti-HIV;  
XX antiviral; antibacterial; antiparkinsonian; antischizoid; antianemic;  
XX antiarthritic; cancer; antirheumatic; hepatoprotective;  
XX antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;  
XX antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;  
XX cardiovascular disorder; neurological disease; nephrotoxic;  
XX gene therapy; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200277186-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 26-MAR-2002; 2002WO-US009188.  
XX  
XX 27-MAR-2001; 2001US-0278650P.  
XX 12-SEP-2001; 2001US-00950082.  
XX 12-SEP-2001; 2001US-00950083.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM;  
XX WPI; 2003-040583/03.  
XX  
XX New human secreted proteins encoded by genes contained in cDNA clones  
XX (e.g. HCCAC19), useful for preventing, treating or diagnosing e.g. AIDS,  
XX multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or  
XX West Nile fever.  
XX  
XX Disclosure; Page 1953-1963; 2423bp; English.  
XX  
XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the  
XX encoded secreted proteins (ABP99470-ABP99872) useful for preventing,  
XX treating or ameliorating medical conditions e.g. by protein or gene

CC therapy. The genes are isolated from a range of human tissues disclosed  
CC in the specification. The nucleic acids, proteins, antibodies and  
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
CC (a) cancer; e.g. breast and ovarian cancer and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.  
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections  
XX

Sequence 41206 BP; 11685 A; 9614 C; 9725 G; 10182 T; 0 U; 0 Other;

	Query Match	20.0%;	Score 400.8;	DB 10;	Length 41206;
	Best Local Similarity	80.7%;	Pred. No. 6.9e-86;		
	Matches 499;	Conservative 0;	Mismatches 102;	Indels 17;	Gaps 2;
QY	461 TAGCTAAAGGATTGTAAATGACCAATCAGCACTCTGTCTAGCTAAAGATTGTAAAT				
Db	36233 TAGCTAAAGGATTGTAAAGGACCAATCAGCACTGTAA-----AAAT				
QY	521 GCAACCAATCAGCACTGTAAATGACCAATCAGCAAGATGTGGCGGGGTCAAAATAG				
Db	36189 GGAACCAATCAGCGCTCTGTAAATGACCAATCAGCAAGATGTGGGTGGGCAAAATAG				
QY	581 GGAGTAAATGAGCCACCCGAGCCAGCACTGGCAACCCACTGGGTCCCTCCACACT				
Db	36129 GGAAATAAAGCTGGCCACCTGCGCGGGGGCAACCCGCTGGGTACTTTACATGCT				
QY	641 GTGGAGCTTTGTTCTTTTCTCTTCAATAATCTTGCTGCTCATTTCTTTGTC				
Db	36069 GTGGAGCTTTGTTCTTTTCTCTTCAATAATCTTGCTGCTCATTTTTGGTC				
QY	701 CACACTACCTTTATGAGCTGTAACTACTGCGAGGGTGTGGCTTCAATCTGAAGT				
Db	36009 TGTACTACCTTTATGAGCTGTAACTACTGCGAGGGTGTGGCTTCAATCTGAAGT				
QY	761 CAACAGACCAAGAACCCACTGGAAGAACAAAGAACTCCGATGTGCTTTAAGAGC				
Db	35949 CAGCAATATCCAGCAACCCCGAGAGAACAAACACTCCCAAGTGCACCTTTAAGAGC				
QY	821 TGTAACTACTGCGAGGCTTGCAAGCTTCACTCCTGAAGTCAAGTGAACCAAAACC				
Db	35889 TGTAACTACTGCGAGGCTTGCAAGCTTCACTCCTGAAGTCAAGTGAACCAAAACC				
QY	881 ACCAAGAGAAAGAACTCTGGAACAACCTGATATCTGAAGAACAACTCAGACAC				
Db	35829 ACCAAGAGAAAGAACTCTGGAAGTACCTGTGAACATCTGAAGAACAAAGCAGACAC				
QY	941 CATCTTCAGAGCTGTAACTCAACCGAAGGGTCTGTGGCTTCAATCTGAAGTCAACA				
Db	35769 CAGCTTTAAGAACTGTCAACACTCACTGCGAGGGTCC-CCAGCTTCATTTTGAAGTCA				
QY	1001 AGACCAAGAACCAACCGAAGAACAAATTCAGACACAGTGAAGAAATCTGTAATTTGA				
Db	35710 AGACCAAGAACCAACCGAAGAACCAATTCAGACACAGTGAAGAAATCTGTAATTTGA				
QY	1061 TCTGTGGCTTCCAGGGTT 1078				
Db	35650 ACCTTCGCTCTCTGGTT 35633				

Search completed: January 22, 2006, 04:09:37  
Job time : 830.081 secs



98 300.6 15.0 39498 3 US-09-949-016-16505 Sequence 16505, A  
C 99 299.8 15.0 128779 3 US-09-497-855A-38 Sequence 38, Appl  
C 100 299.4 15.0 601 3 US-09-949-016-155802 Sequence 155802,

## ALIGNMENTS

## RESULT 1

US-09-078-294-3  
; Sequence 3, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; CURRENT FILING DATE: 1998-05-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 80595  
; TYPE: DNA  
; ORGANISM: Nucleotide sequence of HC-conf1g  
US-09-078-294-3

Query Match 100.0%; Score 2001; DB 3; Length 80595;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGCTTCTGAATCAGACAAATGCTTAAATCCATACCACTCTGAGTTGGG 60  
DB 59000 AAGGCTTCTGAATCAGACAAATGCTTAAATCCATACCACTCTGAGTTGGG 59059  
QY 61 GACATGGCTCTCCCTCTTCTAGGTCCTGACAGCCATCTTGCTAATAGTCGATTTGG 120  
DB 59060 GACATGGCTCTCCCTCTTCTAGGTCCTGACAGCCATCTTGCTAATAGTCGATTTGG 59119  
QY 121 GCCCTGATTTTAACTCTGCTTGTGCAATTTGTTTCTCTAGAGTCGAGGCCATCAAGCT 180  
DB 59120 GCCCTGATTTTAACTCTGCTTGTGCAATTTGTTTCTCTAGAGTCGAGGCCATCAAGCT 59179  
QY 181 ACAGATATCTTAAATGTAACCCCAATGAGCTCAACAACTTCTGCTGAGGACC 240  
DB 59180 ACAGATATCTTAAATGTAACCCCAATGAGCTCAACAACTTCTGCTGAGGACC 59239  
QY 241 CCTGACCGACCGGCTGAGCTTCAATGAGCTTAAAGAGCTCCCTCTGAGGACACTAC 300  
DB 59240 CCTGACCGACCGGCTGAGCTTCAATGAGCTTAAAGAGCTCCCTCTGAGGACACTAC 59299  
QY 301 CACTGCAAGGCGCCCTTCTTCAACCCCTATCCAGACGAAATGACTACAGCGGTCACTGCA 360  
DB 59300 CACTGCAAGGCGCCCTTCTTCAACCCCTATCCAGACGAAATGACTACAGCGGTCACTGCA 59359  
QY 361 AATCCCAACAGACCTGGGCTGCTCTGTTTGAAGGGGAGATTGAAGTGAAGCACTG 420  
DB 59360 AATCCCAACAGACCTGGGCTGCTCTGTTTGAAGGGGAGATTGAAGTGAAGCACTG 59419  
QY 421 GGCTTCTGGGTCAAGTGGGGAATTGGAACCTTTTGTGCTTGAAGTGAAGTGAATG 480  
DB 59420 GGCTTCTGGGTCAAGTGGGGAATTGGAACCTTTTGTGCTTGAAGTGAAGTGAATG 59479  
QY 481 CACCAATCAGACTCTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACTCTGTA 540  
DB 59480 CACCAATCAGACTCTGTGTCTAGCTAAAGATTGTAATGACCAATCAGACTCTGTA 59539  
QY 541 AAATGACCAATCAGACGAGATGTTGGGCGGGCTCAATATAGGAGTAAATCTGGCCACC 600  
DB 59540 AAATGACCAATCAGACGAGATGTTGGGCGGGCTCAATATAGGAGTAAATCTGGCCACC 59599  
QY 601 GAGCGACAGTGGGCAACCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGCTTTTG 660

DB 59600 GAGCGACAGTGGGCAACCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGCTTTTG 59659  
QY 661 CTTTCAACAATTAATCTTGCTGCTGCTCAATCTTTGTGTCCACTACTTTATGAGCTG 720  
DB 59660 CTTTCAACAATTAATCTTGCTGCTGCTCAATCTTTGTGTGTCCACTACTTTATGAGCTG 59719  
QY 721 TAAACCTCAGTGGAGGCTGTGTGCTTCAATCTTGAAGTCAACAGACCAACCACT 780  
DB 59720 TAAACCTCAGTGGAGGCTGTGTGCTTCAATCTTGAAGTCAACAGACCAACCACT 59779  
QY 781 GGAAGGACAAAGAACTCCCAATGTGTGCTCTTTAAGAGCTTTAAGCTCACTGAGGAGC 840  
DB 59780 GGAAGGACAAAGAACTCCCAATGTGTGCTCTTTAAGAGCTTTAAGCTCACTGAGGAGC 59839  
QY 841 TCTGAGCTTCACTCTGTAAGTCAAGTGAAGACCAAAACCCACAGAGGAAAGAACTCTG 900  
DB 59840 TCTGAGCTTCACTCTGTAAGTCAAGTGAAGACCAAAACCCACAGAGGAAAGAACTCTG 59899  
QY 901 GACACACTGAATATCTGAAGGAAACAACTCCAGACACACATCTTTCAGAGCTGTACA 960  
DB 59900 GACACACTGAATATCTGAAGGAAACAACTCCAGACACACATCTTTCAGAGCTGTACA 59959  
QY 961 CTGACCGCAAGGCTGTGTGCTTCAATCTTGAAGTCAAGGAAAGCAAGAACCCACCGGAA 1020  
DB 59960 CTGACCGCAAGGCTGTGTGCTTCAATCTTGAAGTCAAGGAAAGCAAGAACCCACCGGAA 60019  
QY 1021 GGAACAAATTCAGACAGTGAAGTGAATCTGATTTTGTATCTGAGGCTTCCAGAGTTAC 1080  
DB 60020 GGAACAAATTCAGACAGTGAAGTGAATCTGATTTTGTATCTGAGGCTTCCAGAGTTAC 60079  
QY 1081 TCCAGTCAATGAAAGTCTCCATTTGACGCTTAAAGAAACAGAAATGTTTGAAGAGCAC 1140  
DB 60080 TCCAGTCAATGAAAGTCTCCATTTGACGCTTAAAGAAACAGAAATGTTTGAAGAGCAC 60139  
QY 1141 ATGTGGAATTTGTATGACACAGGCTTGAATGACATAGGGCATTTGATTAACCTTA 1200  
DB 60140 ATGTGGAATTTGTATGACACAGGCTTGAATGACATAGGGCATTTGATTAACCTTA 60199  
QY 1201 GCTGGAAGCAGGCGCAGAAATTAATCTAAGGAACAGTTTGTGACAGTATAGT 1260  
DB 60200 GCTGGAAGCAGGCGCAGAAATTAATCTAAGGAACAGTTTGTGACAGTATAGT 60259  
QY 1261 CTTTGCATCTGACACATGTAGATTATCAAGCAATTAATTAAGAAATATATGACAGGTGC 1320  
DB 60260 CTTTGCATCTGACACATGTAGATTATCAAGCAATTAATTAAGAAATATATGACAGGTGC 60319  
QY 1321 GATGCTCATGCTGTATATCCAGACCTTTGGAGGCGCAAGGGGTGTGATCAAGAGTTC 1380  
DB 60320 GATGCTCATGCTGTATATCCAGACCTTTGGAGGCGCAAGGGGTGTGATCAAGAGTTC 60379  
QY 1381 AGCGCTTCAAGACACAGCTGGGCCAATGAGTGAACCCCGTCTCTAATAAATATACAAA 1440  
DB 60380 AGCGCTTCAAGACACAGCTGGGCCAATGAGTGAACCCCGTCTCTAATAAATATACAAA 60439  
QY 1441 ATTAGCTGTGTGTGAGCAGCATCTGTAATCCAGTACTCAGAGGCTGAGGCAAGGG 1500  
DB 60440 ATTAGCTGTGTGTGAGCAGCATCTGTAATCCAGTACTCAGAGGCTGAGGCAAGGG 60499  
QY 1501 AATCTCTTGAACCTTGGAGGCAAGGTTGCAAGTGAAGCCAGATCAACCAACAGACTCCA 1560  
DB 60500 AATCTCTTGAACCTTGGAGGCAAGGTTGCAAGTGAAGCCAGATCAACCAACAGACTCCA 60559  
QY 1561 TCTGAGGTGACAGAGCGAAGCTGTGTCTCAAAAAAAGAAAAAAGAAAGGAAAT 1620  
DB 60560 TCTGAGGTGACAGAGCGAAGCTGTGTCTCAAAAAAAGAAAAAAGAAAGGAAAT 60619  
QY 1621 ATATCAAGATATTTGACAGGTAACTTTATTAACCTTACTATGACCAAGCAATATCA 1680  
DB 60620 ATATCAAGATATTTGACAGGTAACTTTATTAACCTTACTATGACCAAGCAATATCA 60679  
QY 1681 CTAAGTGTTTTACATGAGATTAACTATTAATCTTAACAATAGCCCTATGAGTCAAGTGC 1740

Db	60680	CTAAGTGTTTTAACTAGATTAACTGATTAACTTAACTTAAACAATAGACCTTAATGAGTCAATGCC	60739
Qy	1741	TGTTATTAATCCCACTTTATATGATTAAGAAACTGAAGTAACAAGAGTCAAGTAGAGAA	1800
Db	60740	TGTTATTAATCCCACTTTATATGATTAAGAAACTGAAGTAACAAGAGTCAAGTAGAGAA	60799
Qy	1801	TGGCATTCTGCATTCAGTTTCTAGTTTTTGAAGCACTGTAAACAGAAATCTGTGTGAGAAAT	1860
Db	60800	TGGCATTCTGCATTCAGTTTCTAGTTTTTGAAGCACTGTAAACAGAAATCTGTGTGAGAAAT	60859
Qy	1861	GCTCTAACAAAGATGTGAGTCAAGGGGTTGGAGGTACTGACTCTGAGTTGGCAGTTGGG	1920
Db	60860	GCTCTAACAAAGATGTGAGTCAAGGGGTTGGAGGTACTGACTCTGAGTTGGCAGTTGGG	60919
Qy	1921	ATGGAAGGATGATGAAGAAACAGTTTAAACAAGAGCTGCACTTGGCAACTCTGTGGGA	1980
Db	60920	ATGGAAGGATGATGAAGAAACAGTTTAAACAAGAGCTGCACTTGGCAACTCTGTGGGA	60979
Qy	1981	CTTTGAAGGGTTAAGGGACT	2001
Db	60980	CTTTGAAGGGTTAAGGGACT	61000

RESULT 2  
US-09-078-294-4  
; Sequence 4, Application US/09078294

```

? PARENT INO: 020321-
? GENERAL INFORMATION:
? APPLICANT: Choo, Kong-Hong Andy
? APPLICANT: Du Sart, Desiree
? APPLICANT: Cancilla, Michael R.
? TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
? FILE REFERENCE: Davies Col
? CURRENT APPLICATION NUMBER: US/09/078, 294
? CURRENT FILING DATE: 1998-05-13
? NUMBER OF SEQ ID NOS: 29
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 4
? LENGTH: 80246
? TYPE: DNA
? ORGANISM: Nucleotide sequence of NC-contig
? US-09-078-294-4

```

Query Match	99.9%;	Score 1999.4;	DB 3;	Length 80246;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2000; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY	1	AAAAGCTTCTGAAATGAGACAAATGCCCTTTGAAACTCTTATPACCAACTCTGAGATTGGGC	60
Db	58767	AAAAGCTTCTGAAATGAGACAAATGCCCTTTGAAACTCTTATPACCAACTCTGAGATTGGGC	58826
OY	61	GACATGGCTTCTCCCTTTCTAGGTCCTGTGACAGCATCTTGCTATATGTGGCATTTGG	120
Db	58827	GACATGGCTTCTCCCTTTCTAGGTCCTGTGACAGCATCTTGCTATATGTGGCATTTGG	58886
OY	121	GCCTGTATTTTAACTCTTGGTCAAAATTTGTTTCTCTAGAGTGAAGCCATAGCT	180
Db	58887	GCCTGTATTTTAACTCTTGGTCAAAATTTGTTTCTCTAGAGTGAAGCCATAGCT	58946
OY	181	ACAGATATCTCTCAAAATGTAAACCCCAAAATGACCTCACTAACATCTGTGTGAGAGC	240
Db	58947	ACAGATATCTCTCAAAATGTAAACCCCAAAATGACCTCACTAACATCTGTGTGAGAGC	59006
OY	241	CTTGAGACCGAACCCGCTGGCCCTTTCAATGGCTTAAAGAGCTCCCTCTTGAGAGACATAC	300
Db	59007	CTTGAGACCGAACCCGCTGGCCCTTTCAATGGCTTAAAGAGCTCCCTCTTGAGAGACATAC	59066
OY	301	CACGTGAGAGGGCCCTTCTTACCCCTATACGAGAGAGATGACTACAGCGGTCACTGGCA	360
Db	59067	CACGTGAGAGGGCCCTTCTTACCCCTATACGAGAGAGATGACTACAGCGGTCACTGGCA	59126
OY	361	AATCCCAACAGACACTGGGGTGTCTGTGTTTGAAGGGGGGATTGAGAGGTGAAGCCAGCTG	420

Db	59127	AAATCCCAACAGCAGCTGGGGTGTCTGTTTGAAGGGGGGAATGAGAGTGAAGCCAGCTG	59186
OY	421	GGCTTCTGGGTCAAGTGGGGACTTGGAGACTTTTGTGTCTAGCTAAAGATTTGTAATG	480
Db	59187	GGCTTCTGGGTCAAGTGGGGACTTGGAGACTTTTGTGTCTAGCTAAAGATTTGTAATG	59246
OY	481	CACCAATCAGCACTGTGTCTTAGCTAAAGATTTGAATGCACCAATCAGCACTGTGTA	540
Db	59247	CACCAATCAGCACTGTGTCTTAGCTAAAGATTTGAATGCACCAATCAGCACTGTGTA	59306
OY	541	AAATGACCAATCAGAGATGTGGCGGGGTCAATTAAGGAGTAAAACTGGCCACC	600
Db	59307	AAATGACCAATCAGAGATGTGGCGGGGTCAATTAAGGAGTAAAACTGGCCACC	59366
OY	601	GAGCCAGAGTGGCAACCACTGGGGTCCCTTCCACACTGTGGACCTTGTCTTTTG	660
Db	59367	GAGCCAGAGTGGCAACCACTGGGGTCCCTTCCACACTGTGGACCTTGTCTTTTG	59426
OY	661	CTCTTCAATTAATCTTGTCTGTCTCATTTCTTGTGTCCACTACTTTATAGGTG	720
Db	59427	CTCTTCAATTAATCTTGTGTCTGTCTCATTTCTTGTGTCCACTACTTTATAGGTG	59486
OY	721	TAACTCACTGCGAGGGTCTGTGGCTTACTTCTGAAGTCAACAGCAACCACT	780
Db	59487	TAACTCACTGCGAGGGTCTGTGGCTTACTTCTGAAGTCAACAGCAACCACT	59546
OY	781	GGAAGGAACAAGAACTCCGAGTGTCTGTAAAGAGTGAACCTCACTGGGAAGC	840
Db	59547	GGAAGGAACAAGAACTCCGAGTGTCTGTAAAGAGTGAACCTCACTGGGAAGC	59606
OY	841	TCTGAGCTTCACTCTCGAAGTCAAGAGACCAAAACCCAGAGAGAAACTCTG	900
Db	59607	TCTGAGCTTCACTCTCGAAGTCAAGAGACCAAAACCCAGAGAGAAACTCTG	59666
OY	901	GACACCTGAATATCTGAAGGAACAACTCCAGACACACATCTTTCAGAGCTGAACA	960
Db	59667	GACACCTGAATATCTGAAGGAACAACTCCAGACACACATCTTTCAGAGCTGAACA	59726
OY	961	CTCACCGCAAGGGTCTGTGGCTTCACTTTGAAGTCAAGAACCAAGAACCCAGGAA	1020
Db	59727	CTCACCGCAAGGGTCTGTGGCTTCACTTTGAAGTCAAGAACCAAGAACCCAGGAA	59786
OY	1021	GGAAACAATTCAGACACACAGTAGAAGAACTGTATTTTGTGATCTCGAGCTTCAAGGTTAC	1080
Db	59787	GGAAACAATTCAGACACACAGTAGAAGAACTGTATTTTGTGATCTCGAGCTTCAAGGTTAC	59846
OY	1081	TCCAGTCAATGAAGTCTCATTTGCAGCTTTAAGGAAAACAGAGATGTTTGAAGACAC	1140
Db	59847	TCCAGTCAATGAAGTCTCATTTGCAGCTTTAAGGAAAACAGAGATGTTTGAAGACAC	59906
OY	1141	ATGTGGGAATTTTATATGACACAGGCTTGAATGCACATAGGGATTTCTGATCAAACTTA	1200
Db	59907	ATGTGGGAATTTTATATGACACAGGCTTGAATGCACATAGGGATTTCTGATCAAACTTA	59966
OY	1201	GCTGAAGACAGGGCCAGAAATATATCTTAAGGAACAGTTTTTTGTAGACAGTAGTAGT	1260
Db	59967	GCTGAAGACAGGGCCAGAAATATATCTTAAGGAACAGTTTTTTGTAGACAGTAGTAGT	60026
OY	1261	CTTTGCAATCTGAGACAGTAGTAGATTATCAAGCAATTAATTAAGAAAAATATAGCCAGTGC	1320
Db	60027	CTTTGCAATCTGAGACAGTAGTAGATTATCAAGCAATTAATTAAGAAAAATATAGCCAGTGC	60086
OY	1321	GATGGCTCATGCTGTGTAATCCCAAGCACTTTGGGAGGCCAAAGGGGTGTGATCAACAGTTC	1380
Db	60087	GATGGCTCATGCTGTGTAATCCCAAGCACTTTGGGAGGCCAAAGGGGTGTGATCAACAGTTC	60146
OY	1381	AGGCTTTCAGACACAGCCCTGGCCAAACATGTTGAAACCCCGTCTCTACTAATAAATACAAAA	1440
Db	60147	AGGCTTTCAGACACAGCCCTGGCCAAACATGTTGAAACCCCGTCTCTACTAATAAATACAAAA	60206
OY	1441	ATTAGCTGTGTGTGTGTGCAGCACTCTGTAATCCAGTACTCAGAGGCTGAGGACAGGGG	1500
Db	60207	ATTAGCTGTGTGTGTGTGCAGCACTCTGTAATCCAGTACTCAGAGGCTGAGGACAGGGG	60266

QY	1141	ATGTGGGAATTGTATATGACACAGGCTTGTGATGACATATGGGCGCATTTCTGATCAAACTTA	1200
Db	59847	TCACGTCATTGAAGTCCTCCATTGCGACCCCTTAAGGAAACAGAGATGCTTTGGAGGACAC	59906
QY	59907	ATGTGGGAATTGTATATGACACAGGCTTGTGATGACATATGGGCGCATTTCTGATCAAACTTA	59966
QY	1201	GCTGGAAGCAGGGCCAGGAATATTAATCTTAAGGAAGACACTTTTTGTAGACAGTAGTAGT	1260
Db	59967	GCTGGAAGCAGGGCCAGGAATATTAATCTTAAGGAAGACACTTTTTGTAGACAGTAGTAGT	60026
QY	1261	CTTTGCATCTGAGACATGTAGATTATCAAGCACTTAATAGAAAAATATAGCCAGGTGC	1320
Db	60027	CTTTGCATCTGAGACATGTAGATTATCAAGCACTTAATAGAAAAATATAGCCAGGTGC	60086
QY	1321	GATGCTCATGCTCTGTAAATCCAGCACTTTGGAGGCGCAAGGGGTGTGATCAAGAGTC	1380
Db	60087	GATGCTCATGCTCTGTAAATCCAGCACTTTGGAGGCGCAAGGGGTGTGATCAAGAGTC	60146
QY	1381	AGGCGTTGGAAGCCAGGCTGGCGCAACATGCTGAAACCCCGTCTCTACTAAAAATACAAAA	1440
Db	60147	AGGCGTTGGAAGCCAGGCTGGCGCAACATGCTGAAACCCCGTCTCTACTAAAAATACAAAA	60206
QY	1441	ATTAGCTGTGTGTGTGGCAGCGCATCTGTAAATCCAGTACTCGAAGGCTGAAGCCAGGGG	1500
Db	60207	ATTAGCTGTGTGTGTGGCAGCGCATCTGTAAATCCAGTACTCGAAGGCTGAAGCCAGGGG	60266



QY	1501	AATCTCTTGAACCTTGGAGGCGACAGGTTTGCAGTGAACCAAGATCAACCCACAGACTCCCA	1560
Db	60267	AATCTCTTGAACCTTGGAGGCGACAGGTTTGCAGTGAACCAAGATCAACCCACAGACTCCCA	60326
QY	1561	TCCCGGGGGAACAGAGCGAGACTCGTGTCCAAAAAAGAAAAAAGAAAGGAAAT	1620
Db	60327	TCCCGGGGGAACAGAGCGAGACTCGTGTCCAAAAAAGAAAAAAGAAAGGAAAT	60386
QY	1621	ATAATCAAGATATATTGACAGGTAACTATTATTCAACTACTATGACACGAGCAATACA	1680
Db	60387	ATAATCAAGATATATTGACAGGTAACTATTATTCAACTACTATGACACGAGCAATACA	60446
QY	1681	CTAAGTGTTTTACATGATTAATCACTATTAACTTTAACAAATAGCCCTATGATAGTCACTGC	1740
Db	60447	CTAAGTGTTTTACATGATTAATCACTATTAACTTTAACAAATAGCCCTATGATAGTCACTGC	60506
QY	1741	TGTTATTATCTCCACTTATTATGATAGAAACTGAAATGACGAAAGGTCAAGTATGAGAAA	1800
Db	60507	TGTTATTATCTCCACTTATTATGATAGAAACTGAAATGACGAAAGGTCAAGTATGAGAAA	60566
QY	1801	TGGCAGATGCTGCAGATTTCAGTTTTTGAAGAACTGTACAGAAATCTGGGTGAGAAAT	1860
Db	60567	TGGCAGATGCTGCAGATTTCAGTTTTTGAAGAACTGTACAGAAATCTGGGTGAGAAAT	60626
QY	1861	GCTCTAAACAAGATGTGATCAGAGGTTTGGAGGATCTGAGTCTGAGTTTGGGCACTTGGGG	1920
Db	60627	GCTCTAAACAAGATGTGATCAGAGGTTTGGAGGATCTGAGTCTGAGTTTGGGCACTTGGGG	60686
QY	1921	ATGGAAGATGATGAAGAACAGCTTGACACAGAAAGCTGACACTTGGCACTCTGTGGGA	1980
Db	60687	ATGGAAGATGATGAAGAACAGCTTGACACAGAAAGCTGACACTTGGCACTCTGTGGGA	60746
QY	1981	CCTTGAAGGTTAAGGAGACT	2001
Db	60747	CCTTGAAGGTTAAGGAGACT	60767

```

RESULT 3
US-09-949-002-716
; Sequence 716, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949, 002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 716
; LENGTH: 25589
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25589)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-716

```

	Query Match	Similarity	24.5%	Score	490.6	DB	3	Length	25589
Best Local	Similarity	66.4%	Pred.	No.	4.8e-119				
Matches	868	Conservative	0	Mismatches	379	Indels	60	Gaps	9
OY	323	CCCTATCCAGCAGGAGTACCTACACCGGTCATCGCCCAATCCCAACGACGCTGGGCTG	382						
Db	14437	CACCAATCGGACATCTGTATCTAGCTCAAGGTTGTAAACCAATTCGACACCTCTGTC	14496						
OY	383	TCTGTGTGAGGGGGATTGAGAGGTCAGACGACTGGGCTTCTGGGTCAGGTGGGAC	442						

Db	14497	TAGCTCAGGGTTTGTGATCTGCACCAATCGACACTCTGCATCTAGCTACTGTGTGGGAC	14556
Qy	443	TTGGAGAACTTTTGTG-----TCTAGCTAAAGATTGTAAATGCAACCACTCAGACCTC	495
Db	14557	TTGGAGAACTTTGTGTGTGACACTGTATCTACTATATCTGTGTGGGACGTGGAGAACTTT	14616
Qy	496	TGTGTCTAGCTAAAGATTGTAAATTCACCACTACCACTCTGTAAAA-----543	
Db	14617	TGTTCTTAGGTGTGGGATGGTTAAACCGACCAATCATGTGCTCTGTCAAAACAGACCACTGG	14676
Qy	544	--TGAGCAATTCAGCAGGATGTGGGGGGGTCAAAATAAGAGATAAAAACTGGCCACC	601
Db	14677	GCTTTACCAATTCAGCAGGATGTGGGTGGGGCCAGATTAAAGATTAAAGACAGCTGGCC	14736
Qy	602	AGCCAGAGTGGCAACCACTGGGGTCCCTTTCACACTGTGGAGCTTTGTTCTTTTGGC	661
Db	14737	AGCCAGAGTGGCAACCGTCTGGGGTCCGCTTCACACTGTGGAGCTTTGTTGTTTGGC	14796
Qy	662	TCTTCACAATTAATCTTGTCTGTCTCTCATTTCTTGTGTCCACATACCTTTATGAGCTGT	721
Db	14797	TCCTTTGCAATTAATCTCTGTCTGTCTCTCATCTTTGGGATCCACATGCTTTATGAGCTGT	14856
Qy	722	AACACTCACTGCGAGGGTCTGTGGCTTCATCTCTGAAGTCAAC-AGACCAACCAACCACT	780
Db	14857	GACCTCACTGCGAGGATCTGTCAAGCTTTCATCTCTGAAGCAGCGAGAACCAACAGCCACC	14916
Qy	781	GGAAGAACAAAGAACTCCCGATGTGTGTCTTTAAAGCTGTAACTCACTCACTGCGAAGC	840
Db	14917	AGGAGGAACGAACAATCCAGACGACCGGCTTT-AGAGCTGTAACTCACTGCGAAGG	14975
Qy	841	TCGTGAGCTTCACTCTCTGAAGTCAGTGAGACCAACCAACCAAGAGGAAGAACTCTGG	900
Db	14976	TCTCGAGCTTCACTCTCTGG-AGCCAAGCAGACCAACCAACGAAGAAAGAACTCCA	15034
Qy	901	GACACACCTGAATATCTGAAGAGAACAACTCCAGACACACATCTTTCAGAGCTGTAA	960
Db	15035	AACACATCTGAACTTCAGAAAGAAAAAATCTCCAGCGCGCA-CTTTAAGCTGTAA	15093
Qy	961	CTCACCGGAGGGTCTGTGGCTTCATTTCTTGAAGTCAGCAAGCAAGAACCAACCGGAA	1020
Db	15094	CTCACCGGAGGGTTCACCGCTTCATTTCTTGAAGTCAGTACACCAAGAACCAACCAAT	15153
Qy	1021	GGAACAAATTCAGACACAGTAGAAGATCTGTATTTTGTATCTGTGGCTTCCAGGGTTAC	1080
Db	15154	GCGGACACACCAACCCATTAATCCACTTGAACCACTTGTCTCCGTCAAGGCCCTTAA	15213
Qy	1081	TTCAGTCAATTGAAGTCTCCATTTGCAAGCTTAAGAGAACAGAAATGGTTTGGAGAGCAC	1140
Db	15214	GTTCTTGTGACAAAGGAGAGAAAGTTCAATGTGGGAACTACAGGAGGGCCAAACAG---	15270
Qy	1141	ATGTGGGAATTTATGGAACAGGCTTGAAGTGCATAGGCAATTTCTGATCAAACTTA	1200
Db	15271	-----TTTGTTCATTTTGGTTTACCTTTAACTTTAAT	15306
Qy	1201	GCTGGAACAGGCGCGAGAAATTAATCTTAAGAGAGACAGTTTTTGTAGCAGTAGTATGT	1260
Db	15306	ACAATCTTTAATCACTAGTAAGTAAAGAAAAAGATTAAGAAAAATGT-----AACACAGT	15355
Qy	1261	CTTTGCACTCTGACATGTAGATTATACAGCAATTAATTGAAGAAAAATATAGCAGGTGC	1320
Db	15360	GGTGTGACATATATAGAAAAACACTGTGCTTTAAAAAATTTTGGCCGGCAT	15413
Qy	1321	GATGCTCATGCTGTATAATCCAGACATTTTGGAGGCCAAGGGGTGTGATTCAGAGTCT	1380
Db	15420	GGTGGCTCAACCTGTATATCCAGCAATTTGGAGGGCGAGGTGGCAGATCATTAAGTCT	15473
Qy	1381	AGGCGTTTGAAGCAGGCTGCGCAACATGTGAATCCCGTCTCTACTATAAATAACAAA	1440
Db	15480	AGGAGTTTGAAGCAGGCTGCACTAAACATGGCAAAACCTGTTTTATCTAAAAATACAAA	15533
Qy	1441	ATTTAAGCTGTGTGTGTGGCAGCAATCTGTATCCAG-TTCTAGAGGCTGAGAGCGGG	1499

Db 15540 ATTACCGAGGATGCTGGACATCTGTAATCTCAGCTACTCAGTAGGCTGAGGACGA 15599  
QY 1500 GAATCTCTTGAATCTGGAGGAGGAGGTTGACGTAGGACCAAGATCACACGACACTCC 1559  
Db 15600 GAATCACTTGAACCCAGAGAGGAGGTTGACGTAGGACCAAGATCACGACACTCC 15659  
QY 1560 ATCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA 1606  
Db 15660 AGCTGGGCAACAGAGTAGACTCTGTCTCAAAAAAAAAATAATAATA 15706

RESULT 4  
US-09-949-016-12554/c  
; Sequence 12554, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; LENGTH: 44870  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12554

Query Match 22.9%; Score 457.4; DB 3; Length 44870;  
Best Local Similarity 67.4%; Pred. No. 3,6e-110;  
Matches 781; Conservative 0; Mismatches 354; Indels 23; Gaps 9;

QY 493 CTCTGTGTCTAGTAAAGATTGTAAATGCAACATCAGACTCTGTAA----- 543  
Db 8432 CTTTGTGTCTAGTAAAGATTGTAAATGCAACATCAGACTCTGTAA----- 8373  
QY 544 -----TGACCATCGACGAGATGTGGGGGCTCAATTAAGAGATTAATACTGGCCAC 558  
Db 8372 TCGGTTCTACCAATCAGAGATGTGGGGGCTCAATTAAGAGATTAATACTGGCTGC 8313  
QY 599 CCGAGCCAGCAGTGGCAACCACTCGGGTCCCTTCACACTGTGAAAGCTTTGTTCTTT 658  
Db 8312 CCCAGCCAGCAGTGGCAACCACTCGGGTCCCTTCACACTGTGAAAGCTTTGTTCTTT 8253  
QY 659 TGTCTTTCACATTAATCTTGTCTGTCTCATCTTTGTGTCCACTACTTTATGAGC 718  
Db 8252 CGCTCTTTCACATTAATCTTGTCTGTCTCATCTTTGTGTCCACTACTTTATGAGC 8193  
QY 719 TGTAACTCTACCTGCGAGGCTGTGTGCTTCACTTTCTGAAGTAAAC-AGAACAAGAACCC 777  
Db 8192 TGTAACTCTACCTGCGAGGCTGTGTGCTTCACTTTCTGAAGTAAACCCGAGAGCC 8133  
QY 778 ACTGAAAGAAACAAAGACTCCGATGTGTGCTTTAAGAGTGAACCTGACGCGA 837  
Db 8132 CAGGAGGAAACAAAGACTCCGATGTGTGCTTTAAGAGTGAACCTGACGCGA 8074  
QY 838 AGCTGTGAGCTTCACTCTTGAAGTGAAGTGAACCAAAACCCAGCAAGAAAGAACT 897  
Db 8073 AGCTGTGAGCTTCACTCTTGAAGTGAAGTGAACCAAAACCCAGCAAGAAAGAACT 8015  
QY 898 CTGGAACACCTGAATATCTGAAGAAACAACTCCGAGACACACATCTTTCAGAGCTGA 957  
Db 8014 CCGAAACACCTCTTAATATCTGAAGAAATTAATCTCCAGACGCGCA-CTTTAAGAGCTGTA 7956  
QY 958 ACACTCACCGCAAGGCTGTGTGCTTCACTTTGAAGTGAAGCAAGCAACCAACCG 1017

Db 7955 ACACTCACCGCGAGGCTGTGCGCTTCACTTGAAGTGAAGTGAAGGCAAGATCCACCA 7896  
QY 1018 GAAGAAACAAATTCAGACACAGTAGAAATCTGTATTTTGTATCTGTGCTTCAGAGT 1077  
Db 7895 ATTCTGACACACTTTTACACCACTATTTGTGTCAATCAGGATTTCTTCACTTCCG 7836  
QY 1078 TACTCCAGTCAATGAAGTCTTCATGACGCTTAAGAAACAGAAATGCTTTGAGAG 1137  
Db 7835 CACACATGACAGCTCTCTTCTCTCACTACAGAGACACACTTCTTATCTTACG 7776  
QY 1138 CACATGTGGAAATTTGTATGACCAAGCTTGAGATGACATAGGCAATTTCTGATCAAC 1197  
Db 7775 CTCTGTCTGCAATCGCATGATGATGTTCTTCTAGCTTGAACCATACAGAAAGACAGAC 7716  
QY 1198 CTAGCTGGAAGCAGGCGCAGAAATTAATCTAAGAAAGACATTTTGTAGACAGT 1257  
Db 7715 TCMAATGACAAACCAAAATCAGACCTGACCGCTATTTAGCTCAGGAATGATGAC 7656  
QY 1258 AGTCTTGCATCTGAGACATGTATGATTATCAAGCAATTAATGAATAATATAGCAGG 1317  
Db 7655 AGACTGTAGCTTGAATAAAGCTGACTTCACTTACATCATACCTTTAAGAGGCTCAGG 7596  
QY 1318 TGGATGCTCATGCTCTGTATTCACAGCACTTTGGAGGCAAGGAGTGTGATCAC--G 1375  
Db 7595 CAGGTGCTCACACCTGTATTCAGACACTTTGGAGGCAAGGAGTGTGATCACCTG 7536  
QY 1376 AGTCAAGGCTTGAACCAAGCTGCGCAACATGTGAACCCGCTCTTCTAATAAATA 1435  
Db 7535 AGATTGGAGTTGAGCCACAGCTGACCAATATGAGAAACCCCATCTCTAATAAATA 7476  
QY 1436 CAAATTAAGCTGTGTGTGTGCAAGCATCTGTATCCAG-TACTCAGAGGCTGAGG 1494  
Db 7475 C-AAAAGTAGCGGGGCTGTGCAATGCTGTATCCAGTACTTGGAGGCTGAGG 7417  
QY 1495 CAGGGGAATCTTGTGAATTTGGAGGCAAGTGTGAGTGAAGCAAGATCAACACAGC 1554  
Db 7416 CAGGAATATCGTTTAAACCGGAGGCAAGGTTGAGTAGGCCACATGTGCTTTC 7357  
QY 1555 ACTCATCTCTGTGAC-AGAGCAGACTCTGTCTCAAAAAAAAAAAAAAAAAAGAAA 1613  
Db 7356 ACTCCAGCTGGGCAACAGAGTGAATCTCCGCTCAAAACAAACGAAACAAAAA 7297  
QY 1614 GGAATAATATCAAGAA 1631  
Db 7296 CCACACACCTTAACCA 7279

RESULT 5  
US-09-949-016-16349/c  
; Sequence 16349, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; LENGTH: 44881  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16349

Query Match 22.8%; Score 457.4; DB 3; Length 44881;  
 Best Local Similarity 67.4%; Pred. No. 3.6e-110;  
 Matches 781; Conservative 0; Mismatches 354; Indels 23; Gaps 9;

493 CTCTGCTAGTAAGATGTTGAATGACCAATCAGACCTCTGTAAA-----543  
 DB 8439 CTTTGTCAGCTCAGGATGTGAACGACCAATCAGACCTCTGTCAAAACAGACCAC 8380  
 544 -----TGACCAATCAGCAGATGTGCGCGGGTCAAAATGAAGATAAAACCTGGCCAC 598  
 DB 8379 TGGGTTCTACCATCAGCAGGATGTGGGTGGGGCCAGATAGAGATTAAGAGCTGGCTGC 8320  
 599 CCGAGCCAGCAGTGGCAACCCACTCGGGTCCCTTCACACTGTGAAGCTGTTCTTT 658  
 DB 8319 CCAGCCAGCAGTGGCAACCCGCTTAGTCCCTTCACACTGTGAAGCTGTTCTTT 8260  
 659 TGGCTTTCAAAATATCTTGTCTGTCTCATCTTGTGTCCACACTACCTTATGAGC 718  
 DB 8259 CGCTCTTGAATAATCTTGTCTACTGTCTACTCTTGGGTCCACGCTGTTTATGAGC 8200  
 719 TGTAACTCACTGCGAGGGTCTGTGGCTTCATTCTGAAGTCAAC-AGACCAAGAACCC 777  
 DB 8199 TGTAACTCACTGCGCAAGGTCTGCACTTCACTCTAAAGCCAGCAAGCCGAGGCC 8140  
 778 ACTGAAGAAACAAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCACTGCGA 837  
 DB 8139 CCAGGAGAAACAAAGAACTCCAGACGCGCTGCC-TTAAAGCTGTAACTCACTGCGCA 8081  
 838 AGCTTCGAGCTTCACTCTCTGAAGTCACTGACGACCAACCCAGCAAGAAAGAACT 897  
 DB 8080 AGGTCGAGCTTCACTCTCTG-AGCAGGACGACCAAGAAAGAAAGAAAGT 8022  
 898 CTGACACACCTGAATATCTGAAGAAACAACTCCAGACACACCATCTTTCAGAGCTGA 957  
 DB 8021 CCGAAACATCTCTAATCTGAAGAAATTAATCCAGACACGCCA-CTTAAAGCTGTGA 7963  
 958 ACACTCAACCGCAAGGCTGTGGCTTCATTCTTGAAGTCAAGAACCAAGAAAGAAAGT 1017  
 DB 7962 ACACTCAACCGCAAGGCTGTGGCTTCATTCTTGAAGTCAAGAACCAAGAAAGT 7903  
 1018 GAAGGAACAAATTCAGACACAGTGAAGAAATCTGTATTTTGTGATCTGTGGCTTCAGAGT 1077  
 DB 7902 ATTCTGACACACTTTTACACACTATTTGTGCAATCAGGCTTCTCTCACTTCG 7843  
 1078 TACTCCAGTATGAAGTCTTCATTTGAGCTTGAAGAAACAGAAATGTTGAGAGAG 1137  
 DB 7842 CACACAAATGACGCTCTCTCTCTGACTACAGACACACATTTCTTATACCTACG 7783  
 1138 CACATGTGAATTTGATATGACACAGGCTTGAAGTGAATGAGGCAATTTCTGATCAAC 1197  
 DB 7782 CTCTGCTCAGATGCGATGCTTCTTCTTGAAGTGAAGGCAATTTCTGATCAAC 7723  
 1198 CTGAGCTGAAGCAGAGGCGCAAGAAATTAATCTGAAGAAAGAGAGCTTTTGTGAAGAGT 1257  
 DB 7722 TCAAAATGACAAAGCCAAATCAGAACTGTAAGCGCTTTTGAAGTCAAGAACTGAGTAC 7663  
 1258 AGCTTTGATGAGACATGATGATTAATCAAGCAATTAATTAAGAAAATATATGACGAG 1317  
 DB 7662 AGAACTGTAAGTGAAGAAAGCTGATCTTACCTTTACATCATCTTTAAGAGGACAG 7603  
 1318 TGGATGCTCATGCTCTGTAATCCACACATTTTGGAGGCGCAAGGGTGTGATCAC--G 1375  
 DB 7602 CAGGCTGGCTCAACCTGTATCCGACACTTTGGAGGCGCAAGGGTGGGAGATCACTG 7543  
 1376 AGGTCAGGCTTGAAGACAGGCTGCGCAACATGTAAGAAACCCGCTCTTCTAATAAATA 1435  
 DB 7542 AGATTGGAGATTGAGGCCAGCGCTGACCAATATGAGAAACCCCATCTCTAATAAATA 7483  
 1436 CAAAAATTAAGCTGTGTGTGAGGACGATCTGTAATCCAG--TACTCAGAGGCTGAG 1494  
 DB 7482 C-AAAAATTAAGCTGTGTGTGAGGACATGCTGTAATCCAGCTACTTGTGAGGCTGAG 7424  
 1495 CAGGGAATCTCTTGAATTTGGAGGACAGAGTTGCAGTGAGCAAGATCACACAGC 1554

DB 7423 CAGGGAATCTGCTTAAACCGGAGGCAAGGTTGATGAGCCCACTGCGCACTTGC 7364  
 1555 ACTCCATCTGGGTGAC-AGAGCGAGACTCTGTCAAAAAAAAAAAAAAAAAAGAAA 1613  
 DB 7363 ACTCCAGCTGGGCAACAGAGTGAATCTCGCTCAAAAACAAAGCAAAAAA 7304  
 1614 GGAATATATATCAAGAA 1631  
 DB 7303 CCACAAACCTTAACCA 7286

RESULT 6  
 US-08-691-563C-46  
 ; Sequence 46, Application US/08691563C  
 ; Patent No. 6001987  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HERVE PERRON  
 ; APPLICANT: Frederic BESEME  
 ; APPLICANT: Frederic BEDIN  
 ; APPLICANT: Glaucia PARANHOS-BACCALA  
 ; APPLICANT: Florence KOMURIAN-PRADEL  
 ; APPLICANT: Colette JOLIVET  
 ; APPLICANT: Bernard MANDRAND  
 ; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
 ; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC  
 ; NUMBER OF SEQUENCES: 92  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Oliff & Bertridge  
 ; STREET: 700 South Washington Street, Suite 300  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22314  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/691,563C  
 ; FILING DATE: 02-AUG-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Bertridge, William P.  
 ; REGISTRATION NUMBER: 30,024  
 ; REFERENCE/DOCKET NUMBER: MPB 38588  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-836-6400  
 ; TELEFAX: 703-836-2787  
 ; INFORMATION FOR SEQ ID NO: 46:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1859 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; US-08-691-563C-46

Query Match 22.6%; Score 452.6; DB 3; Length 1859;  
 Best Local Similarity 91.6%; Pred. No. 1.6e-109;  
 Matches 490; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

1 AAAGGCTTGAATGAGACCAATGACCTTCAAACTCTTATACCACTCTGAGTTGGC 60  
 DB 1325 AAAGGCTTGAATGAGACCAATGACCTTCAAACTCTTATACCACTCTGAGTTGGC 1384  
 61 GACATGCTTCTCCCTTTCTAGGCTCTGTGACAGCATCTTCTATAGTGCATTGG 120  
 DB 1385 AACATGTTTCTTCCCTTTATGTCACATGCTGCATCTTCTATTAATGCGCTTGG 1444  
 121 GCCCTGATTTTAACTCTTGGTCAATTTGTTTCTCTAGAGTGAAGGCAATCAAGCT 180

Db 1445 GCCCTGTAATTTTAACTCCCTTGTCAAATTTGTTCTTACAGATGAGCCATCAAGCT 1504  
Qy 181 ACAGATGATCTTCAATATGTAACCCCAATAGAGTCACTTAAGTCTGAGGATC 240  
Db 1505 ACAGATGATCTTCAATATGTAACCCCAATAGAGTCACTTAAGTCTGAGGATC 1564  
Qy 241 CCTGACGACGACCGCTGAGGCTTTCAATGAGCTCCCTCTGAGAGCACTAC 300  
Db 1565 CCTAGACCAACCCCTGAGGCTTTCACTGAGCTTAAGAGTTCCTCTGAGAGCACTAC 1624  
Qy 301 CACTGACGAGGCTCCCTTTTACCCCTATCCAGAGAGTGTAGAGCTGATATGGCC 1684  
Db 1625 CACTGACGAGGCTCCCTTTTACCCCTATCCAGAGAGTGTAGAGCTGATATGGCC 1684  
Qy 360 AATCCCAACAGAGCTGAGGCTGCTGTTTGAAGGAGGATGAGAGGAGCAAGCT 419  
Db 1685 AATCCCAACAGAGCTGAGGCTGCTGTTTGAAGGAGGATGAGAGGAGCAAGCT 1744  
Qy 420 GGGCTTCTGGGTGAGGTGAGGACTTGAGAACTTTGTCTAGCTAAAGATTGTAAT 479  
Db 1745 GAGCTTCTGGGTGAGGTGAGGACTTGAGAACTTTGTCTAGCTAAAGATTGTAAT 1804  
Qy 480 GCAACATCAGTCTGTGTCTAGCTAAAGATTGTAATGACCAATCAGCAGC 534  
Db 1805 GCAACATCAGTCTGTGTCTAGCTAAAGATTGTAATGACCAATCAGCAGC 1859

RESULT 7  
US-09-374-766-46  
; Sequence 46, Application US/09374766  
; Patent No. 6579526  
; GENERAL INFORMATION:  
; APPLICANT: HERVE PERRON  
; APPLICANT: FREDERIC BESEME  
; APPLICANT: GLAUCIA PARANHOS-BACCALA  
; APPLICANT: FLORENCE KOMURIAN-PRADEL  
; APPLICANT: COLETTE JOLIVET  
; APPLICANT: BERNARD MANDRAND  
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC  
; NUMBER OF SEQUENCES: 92  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Oliff & Berridge  
; STREET: 700 South Washington Street, Suite 300  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22314  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/374,766  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/691,563  
; FILING DATE: 02-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berridge, William P.  
; REGISTRATION NUMBER: 30,024  
; REFERENCE/DOCKET NUMBER: WPB 38588  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6400  
; TELEFAX: 703-836-2787  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1859 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; US-09-374-766-46  
Query Match 22.6%; Score 452.6; DB 3; Length 1859;  
Best Local Similarity 91.6%; Pred. No. 1,6e-109;  
Matches 490; Conservative 0; Mismatches 44; Indels 1; Gaps 1;  
Qy 1 AAAGCTTGAAGTAAGCAATGCTTCAAACTCTTTATACCACTTGTGAGTTGGG 60  
Db 1325 AAAGCTTCTGAATATGACACAGCTTCAAACTCTTTATACCACTTGTGAGTTGGG 1384  
Qy 61 GACATGCTTCTCCCTTTCTAGTCTGTGACAGCATCTTGTATATAGTCGATTTGG 120  
Db 1385 AACATGTTTCTCCCTTTCTAGTCTGTGACAGCATCTTGTATATAGTCGATTTGG 1444  
Qy 121 GCCCTGTAATTTTAACTCTTGTGCAATTTGTTTCTCTAGAGTGAAGCCATCAAGCT 180  
Db 1445 GCCCTGTAATTTTAACTCTTGTGCAATTTGTTTCTCTAGAGTGAAGCCATCAAGCT 1504  
Qy 181 ACAGATGATCTTCAATATGTAACCCCAATAGAGTCACTTAAGTCTGAGAGCAGC 240  
Db 1505 ACAGATGATCTTCAATATGTAACCCCAATAGAGTCACTTAAGTCTGAGAGCAGC 1564  
Qy 241 CCTGACGACGACCGCTGAGGCTTTCAATGAGCTCCCTCTGAGAGCACTAC 300  
Db 1565 CACTGACGAGGCTCCCTTTTACCCCTATCCAGAGAGTGTAGAGCTGATATGGCC 1624  
Qy 301 CACTGACGAGGCTCCCTTTTACCCCTATCCAGAGAGTGTAGAGCTGATATGGCC 359  
Db 1625 CACTGACGAGGCTCCCTTTTACCCCTATCCAGAGAGTGTAGAGCTGATATGGCC 1684  
Qy 360 AATCCCAACAGAGCTGAGGCTGCTGTTTGAAGGAGGATGAGAGGAGCAAGCT 419  
Db 1685 AATCCCAACAGAGCTGAGGCTGCTGTTTGAAGGAGGATGAGAGGAGCAAGCT 1744  
Qy 420 GGGCTTCTGGGTGAGGTGAGGACTTGAGAACTTTGTCTAGCTAAAGATTGTAAT 479  
Db 1745 GAGCTTCTGGGTGAGGTGAGGACTTGAGAACTTTGTCTAGCTAAAGATTGTAAT 1804  
Qy 480 GCAACATCAGTCTGTGTCTAGCTAAAGATTGTAATGACCAATCAGCAGC 534  
Db 1805 GCAACATCAGTCTGTGTCTAGCTAAAGATTGTAATGACCAATCAGCAGC 1859

RESULT 8  
US-08-979-847B-42  
; Sequence 42, Application US/08979847B  
; Patent No. 6582703  
; GENERAL INFORMATION:  
; APPLICANT: PERRON, HERVE  
; APPLICANT: BESEME, FREDERIC  
; APPLICANT: PARANHOS-BACCALA, GLAUCIA  
; APPLICANT: KOMURIAN-PRADEL, FLORENCE  
; APPLICANT: JOLIVET-REYNAUD, COLETTE  
; APPLICANT: MANDRAND, BERNARD  
; APPLICANT: GARSON, JEREMY  
; APPLICANT: TURE, PHILIP  
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC  
; NUMBER OF SEQUENCES: 210  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Oliff & Berridge, PLC  
; STREET: P.O. BOX 19928  
; CITY: ALEXANDRIA  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```
/
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/979,844B
/ FILING DATE: 26-Nov-6582703-1997
/ CLASSIFICATION: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BERRIDGE, WILLIAM P.
/ REGISTRATION NUMBER: 30,024
/ REFERENCE/DOCKET NUMBER: WPB 39046A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 703-836-6400
/ TELEFAX: 703-836-2787
/ INFORMATION FOR SEQ ID NO: 42:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1859 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-979-847B-42

Query Match      22.6%; Score 452.6; DB 3; Length 1859;
Best Local Similarity 91.6%; Pred. No. 1.6e-109;
Matches 490; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 1 AAGGCTTCTGAATCAGACAAATGCTTAACTCTTAATACCACTCTGGAGTTGGG 60
DB 1325 AAGGCTTCTGAATCAGACAAATGCTTAACTCTTAATACCACTCTGGAGTTGGG 1384
QY 61 GACATGAGCTTCTCCCTCTTCTAGAGTCTGAGAGCCATCTTGTCTAATATGACATTTGG 120
DB 1385 AACATGCTTCTCCCTCTTCTAGAGTCTGAGAGCCATCTTGTCTAATATGACATTTGG 1444
QY 121 GCCCTGATTTTAACTCTTGTGCAATTTGTTCTCTAGAGTCTGAGAGCCATCAAGCT 180
DB 1445 GCCCTGATTTTAACTCTTGTGCAATTTGTTCTCTAGAGTCTGAGAGCCATCAAGCT 1504
QY 181 ACAGATGATTTTAACTCTTGTGCAATTTGTTCTCTAGAGTCTGAGAGCCATCAAGCT 240
DB 1505 ACAGATGATTTTAACTCTTGTGCAATTTGTTCTCTAGAGTCTGAGAGCCATCAAGCT 1564
QY 241 CCGTGAACGAGCCGCTGAGCTTCTCAATGAGCTTAAAGCTCCCTCTGAGAGCACTAC 300
DB 1565 CCGTGAACGAGCCGCTGAGCTTCTCAATGAGCTTAAAGCTCCCTCTGAGAGCACTAC 1624
QY 301 CACTGAGAGGCCCTTCTTCAACCCCTATCCAGAGAGTACTACAGCGGTGATG-CC 359
DB 1625 CACTGAGAGGCCCTTCTTCAACCCCTATCCAGAGAGTACTACAGCGGTGATG-CC 1684
QY 360 AATTCCTCAACAGACGCTGGGGGTGCTCTGTTTGAAGGGGGGATGAGAGCCAGCT 419
DB 1685 AATTCCTCAACAGACGCTGGGGGTGCTCTGTTTGAAGGGGGGATGAGAGCCAGCT 1744
QY 420 GGGCTTCTGGGGTCAAGTGGGGGACTTGGAGAACTTTGTGTAGCTAAAGGATTTGTAAT 479
DB 1745 GGGCTTCTGGGGTCAAGTGGGGGACTTGGAGAACTTTGTGTAGCTAAAGGATTTGTAAT 1804
QY 480 GCAACATCAACACTCTGTGTCTAGCTAAAGGATTTGTAATGCAATCAAGCAC 534
DB 1805 GCAACATCAACACTCTGTGTCTAGCTAAAGGATTTGTAATGCAATCAAGCAC 1859

RESULT 9
US-09-949-016-15316
; Sequence 15316, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
```

```
/
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 15316
/ LENGTH: 144322
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)..(144322)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15316

Query Match      21.8%; Score 436.6; DB 3; Length 144322;
Best Local Similarity 88.9%; Pred. No. 1.9e-104;
Matches 506; Conservative 0; Mismatches 59; Indels 4; Gaps 3;

QY 476 AATGACCAATCAGACTCTGTGTCTAGCTTAAAGGATTTGTAATGACCAATCAGACT 535
DB 15433 AATGACCAATCAGACTCTGTGTCTAGCTTAAAGGATTTGTAATGACCAATCAGACT 15492
QY 536 CTGTAAATATGAGCAATCAGAGAGTGTGGGGGCTCAATTAAGGAGATTAATACTGGC 595
DB 15493 CTGTAAATATGAGCAATCAGAGAGTGTGGGGGCTCAATTAAGGAGATTAATACTGGC 15552
QY 596 CACCCGAGCAGAG-TCGCAACCACTGAGTCCCTTCACTGTGGAAGCTTTGT 654
DB 15553 CACCCGAGCAGAG-TCGCAACCACTGAGTCCCTTCACTGTGGAAGCTTTGT 15612
QY 655 CTTTGTCTCTCAATTAATCTTGTGCTGCTCAATTTCTTGTGTCACTACCTTTAT 714
DB 15613 CTTTGTCTCTCAATTAATCTTGTGCTGCTCAATTTCTTGTGTCACTACCTTTAT 15670
QY 715 GAGCTTAACCTCACTGAGAGGCTGTGGCTTAATCTGTAAGTCAACA-GAGCAGCA 773
DB 15671 GAGCTTAACCTCACTGAGAGGCTGTGGCTTAATCTGTAAGTCAACA-GAGCAGCA 15730
QY 774 ACCCACTGAAGAGCAAAAGAACTCCGATGTGCTGCTTTAAGACTGTAACCTCACT 833
DB 15731 ACCCACTGAAGAGCAAAAGAACTCCGATGTGCTGCTTTAAGACTGTAACCTCACT 15790
QY 834 GCGAAGCTTGCAGCTTCACTCTGAAGTCAAGTGAACAACAACCACTGGAAGAGA 893
DB 15791 GCGAAGCTTGCAGCTTCACTCTGAAGTCAAGTGAACAACAACCACTGGAAGAGA 15850
QY 894 AACTCTGAGACACCTGTAATTTCTGAAGAGAAATCTCCAGACACCACTTTCAAGC 953
DB 15851 AACTCTGAGACACCTGTAATTTCTGAAGAGAAATCTCCAGACACCACTTTCAAGC 15910
QY 954 TGTAACTACACCGCAGAGGCTGTGAGCTTCTTTTGAAGTCAAGAGCAAGAACCC 1013
DB 15911 TGTAACTACACCGCAGAGGCTGTGAGCTTCTTTTGAAGTCAAGAGCAAGAACCC 15970
QY 1014 ACCGGAAGAAATTTCCAGACAGTA 1042
DB 15971 ACCGGAAGAAATTTCCAGACAGTA 15999

RESULT 10
US-09-949-016-17012/c
; Sequence 17012, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17012  
LENGTH: 89625  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(89625)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-17012

Query Match 21.8%; Score 435.8; DB 3; Length 89625;  
Best Local Similarity 66.4%; Pred. No. 2.5e-104;  
Matches 872; Conservative 0; Mismatches 287; Indels 154; Gaps 11;

QY 431 TCAGTGGGGAGCTTGGAGAACTTTGTCTAGCTAAAGATTGTAAATGCAATCAG 490  
DB 38298 TCAGTGGGGAGGTTGGAGAACTTTGTCTAGCTAGGATTTGTAAATGCAATCAG 38239  
QY 491 CACTGTGTCTAGCTAAAGATTGTAAATGCAATCAGCTCTGTAAATGCAACA 550  
DB 38238 CACCTTGACA-----AAACAGCAATCAGCTCTGTAAATGCAACA 38195  
QY 551 ATCAGCAGATGTGGGGGGGCTCAATTAAGGAGTAAAACTGGCCACCCGAGCCAGAG 610  
DB 38194 ATCAGCAGATGTGGGGGGGCTCAATTAAGGAGTAAAACTGGCCAGCCAGAG 38135  
QY 611 TGGCAACCACTGGGGTCCCTTCCACACTGTGGAAGCTTTCTTTTGTCTTCAACA 670  
DB 38134 TGGCAACCTGGGGTCCCTTCCACACTGTGGAAGCTTTCTTTTGTCTTCAACA 38075  
QY 671 TAAATCTGTCTGTCTCAATTTCTTGTCTCACTAATCTTAAAGCTTAACACTCAG 720  
DB 38074 TAAATCTGTCTGTCTCAATTTCTTGTCTCACTAATCTTAAAGCTTAACACTCAG 38015  
QY 731 TGGAGGGTCTGTGGCTTCAATCTGAAGTCAAC-AGACCAAGCAACCCACTGAGAGAAC 789  
DB 38014 TGGAGGGTCTGTGGCTTCAATCTGAAGTCAAC-AGACCAAGCAACCCACTGAGAGAAC 37955  
QY 790 AAAAGAACTCCCAATGTGTCTTAAAGCTTAACACTCTGGAAGCTCTGCAGCT 849  
DB 37954 GAAACAATCTCCCAATGTGTCTTAAAGCTTAACACTCTGGAAGCTCTGCAGCT 37896  
QY 850 TCACCTCTGAAGTCAAGTGAAGCAACCAACCAAGGAGAACTCTGAGCAACACT 909  
DB 37895 TCACCTCTGAAGTCAAGTGAAGCAACCAACCAAGGAGAACTCTGAGCAACACT 37837  
QY 910 GAATATCTGAAGCAACCACTCCAGACACCACTTTTGAAGCTTAACTACCTCAGCAG 969  
DB 37836 GAATATCTGAAGCAACCACTCCAGACACCACTTTTGAAGCTTAACTACCTCAGCAG 37777  
QY 970 AGGGTCTGTGGCTTCAATCTTGAAGTCAAGCAACCAAGCCAGC-----GGAAG 1022  
DB 37776 AGGGTCTGTGGCTTCAATCTTGAAGTCAAGCAACCAAGCCAGC-----GGAAG 37717  
QY 1023 AACCAATCTCAGACAGTGAAGAACTGTATTTTGTATGTGGCTT-----1070  
DB 37716 CTAACATGTTTTCAGACCTGAAATATCTTTTGTATTTTGTATTTTGTATTTTGTATTTT 37657  
QY 1071 -----CAGGGTACTCAGATCAATGAAGTCTCCATTCAGCTTAAGGAAA 1117  
DB 37656 TTTGAGATCTGAGCTCAGATTCACCAATGTTATTGGCTTAAGCAGACAGTTATGCAAG 37597

QY 1118 CAGAGATGTTTGGAGGACATGTGGGAATTTGTAT-----1156  
DB 37596 CTGTCAATTTGCCAAAGCTCAATAGCTGTGTAACAGTATGATCTTCCAGTGGCTATGGCAT 37537  
QY 1157 ---GACCAAGCTTGAAGTCAATAGGCAATTTCTGATCAAACTTACCTGAGAGAGG 1213  
DB 37536 TGGTACACATGTTGAGTGTGATGATTTAAATGCTTTTAAAACTTATGATTTAGCACTG 37477  
QY 1214 CCAGGAATATATA-TCTAAGGAAGACATTTTGTATGACAGTATGCTTTGCAATCTG 1271  
DB 37476 AGAAGAAAGTTATCTCAGAGAAAGCAAAAGCAGAGCATGTGATTTGTATGTT 37417  
QY 1272 AGACATGATGATTTATCA-----1288  
DB 37416 AGTGAACCAATATTTTATTAAGCTATTAACAAAGCAATGTTTCTTCCAAAGC 37357  
QY 1289 -----AGCAATTAATTAAGAAATATATGACAGG 1317  
DB 37356 AACCAACAGCCCTTAATGAAGCCCAATAGAGAAACCTTTAAAGATTAAGCTG 37297  
QY 1318 TGGATGCTCTCAATGCTGTATATCCAGCACTTTGGAGCCCAAGGCTGTGATCAAC--G 1375  
DB 37296 GCTGTGGCTTCAACCTGTATATCCAGCACTTTGGAGGCTGTGATGATCAACCTG 37237  
QY 1376 AGGTCAAGGCTTGAAGCAGCCTGAGCAACATGTGAACCCCGTCTACTTAAATA 1435  
DB 37236 AGGTCAAGGCTTGAAGCAGCCTGAGCAACATGTGAACCCCGTCTACTTAAATA 37177  
QY 1436 CAAAATTAAGCTGTGTGTGTGAGCAGCATCTGTATATCCAG-TACTCAGAGGCTGAGG 1494  
DB 37176 CAAAATTAAGCTGTGTGTGTGAGCAGCATCTGTATATCCAGTATCTCAGAGGCTGAGG 37117  
QY 1495 CAGGGGAATCTCTTGAACCTTGGAGCAGAGGTTGAGAGCCAAAGATCAACCAAGC 1554  
DB 37116 CAGGGGAATCTCTTGAACCTTGGAGCAGAGGTTGAGAGCCAAAGATCAACCAAGC 37057  
QY 1555 ACTCCATCTCGGGTGAAGAGGAGCATCTGTCTCAAAAAAATTTTTTTTTT 1607  
DB 37056 ACTCCATCTCGGGTGAAGAGGAGCATCTGTCTCAAAAAAATTTTTTTTTT 37004

RESULT 11  
US-09-949-016-13173/C  
Sequence 13173, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13173  
LENGTH: 360470  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-13173

Query Match 21.6%; Score 432.8; DB 3; Length 360470;  
Best Local Similarity 84.8%; Pred. No. 2.8e-103;  
Matches 519; Conservative 0; Mismatches 67; Indels 26; Gaps 2;  
QY 431 TCAGTGGGGAGCTTGGAGAACTTTGTCTAGCTAAAGATTGTAAATGCAATCAG 490  
DB 233842 TCAGTGGGGAGCTTGGAGAACTTTGTCTAGCTAAAGATTGTAAATGCAATCAG 233783



QY	491	CACTCTGTGTCTAGACTTAAGATTTGTAAATGCAACAATACGA-----	533
Db	233782	CACCTCTGTGTCTAGACTCAAGGTTTGTAAACGACCAATAGCAACTGTGAAAAACGACC	233723
QY	534	-----CTCGTAAATATGATGACCAATCAGAGSAGTGTGGCGGGGTCAAAATPAGGAGT	585
Db	233722	AATCAGCTCTCTGTATAAATGGGSCCAATTGGCAGAGATGTGGGTGAGGTCAGATPAGGAAAT	233663
QY	586	AAAACTGGCCACCCCGAGCCAGCAAGTGGCAACCCACTCGGGTCCCTTCCACACTGTGGA	645
Db	233662	AAAAACAGAGCTCCCAAGCCAGACAGCAACCTGCTGGGGTCCCTCCATGCTGTGGA	233603
QY	646	AGCTTGTCTTTTGGTCTCAACAATAAATCTTGGCTGCTCATCTTGTGTCCACAC	705
Db	233602	AGCTTGTCTTTTCACTCTTCAACAATAAATCTTGTGTCTGCTCATCTTGTGTCCACAC	233543
QY	706	TACCTTTATGAGCTGTAACTCACTCACTGCGAGGGTCTGTGGCTTCAATCTCTGAAGTCAAC-	764
Db	233542	TGCTTTTAAAGACTGTAAACACTCAAGGTGAAGGTCTGCAAGCTTCACTCTCTGAAGCCAGCA	233483
QY	765	AGACCAACGAACCACTGGAAGAAACAAGAATCTCCGATGTGCTGCTTTAAGAGCTGTA	824
Db	233482	AGACCAACGAACCACTGGAAGAAACAAGAATCTCCGATGTGCTGCTTTAAGAGCTGTG	233423
QY	825	ACACTCACTGCGAGAGCTGCAAGCTTCACTCCCTGGAAGTCAAGTGAACAACAACCCACCA	884
Db	233422	ACACTCACTGCGAGAGCTGCAAGCTTCACTCCCTGGAAGTCAAGTGAACAACAACCCACCA	233363
QY	885	GAAAGAAAGAACTCTTGACACACCTGSAATATCTGAAGAAACAATCTCCAGACACACATC	944
Db	233362	GAAAGAAAGAAATCTTGACACACCTGSAATATCTGAAGAAACAATCTCCAGACACACATC	233303
QY	945	TTTTCAGAGCTGTAACTCAACCGCAAGGGTCTGTGGCTTCAATCTTGAATCAGCAAGAC	1004
Db	233302	TTTTCAGAGCTGTAACTCAACCGCAAGGGTCTGTGGCTTCAATCTTGAATCAGCAAGAC	233243
QY	1005	CAAGAACCCACCC 1016	
Db	233242	CAAGAACCTCACCC 233231	
RESULT 12			
US-09-949-016-16241/C			
Sequence 16241, Application US/09949016			
Patent No. 6812339			
GENERAL INFORMATION:			
APPLICANT: VENTER, J. Craig et al.			
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED			
FILE REFERENCE: CL001307			
CURRENT APPLICATION NUMBER: US/09/949, 016			
CURRENT FILING DATE: 2000-04-14			
PRIOR APPLICATION NUMBER: 60/241,755			
PRIOR FILING DATE: 2000-10-20			
PRIOR APPLICATION NUMBER: 60/237,768			
PRIOR FILING DATE: 2000-10-03			
PRIOR APPLICATION NUMBER: 60/231,498			
PRIOR FILING DATE: 2000-09-08			
NUMBER OF SEQ ID NOS: 207012			
SOFTWARE: FastSeq for Windows Version 4.0			
SEQ ID NO 16241			
LENGTH: 84132			
TYPE: DNA			
ORGANISM: Human			
FEATURE:			
NAME/KEY: misc_feature			
LOCATION: (1)...(84132)			
OTHER INFORMATION: n = A,T,C or G			
US-09-949-016-16241			
Query Match 21.1%; Score 421.6; DB 3; Length 84132;			
Best Local Similarity 68.4%; Pred. No. 1,4e-100;			

	Matches	812;	Conservative	0;	Mismatches	214;	Indels	162;	Gaps	10;
QY	437	GGGGA	CTTGAGAACTTTTGTGTCTAGCTAAAGATTGTAATGCAACCATCAGACTCT	496						
Db	21963	GAGAA	CATGAGAACTTTTGTGTCTAGCTCAGGGATTGTAACGCAACCAATCAGACCCCT	21904						
QY	497	GTGT	-----CTAGCTAAAGATTGTAATGCAACCAATCAGACTCTGTAAATATGAC	548						
Db	21903	GT	MAAACCGAACCAATCAACTCTGTGTAAACAGACCAATCAGCTCTGTAAATATGAC	21844						
QY	549	CAAT	CAGCAGAGATGTGGCGGGGTCAAAATTAAGGAGTAAAACTGGCCACCCGAGCAGC	608						
Db	21843	CAAT	CAGCAGAGATGTGCGTGGGGCCAGATAGAGATAATAAAACAGCGCTGTGAGCCAGC	21784						
QY	609	AGTGG	CAACCCACTCGGGTCCCTTCCTCACTGTGGAAAGCTTTGTTCTTTTGCTCTTAC	668						
Db	21783	AGTGG	CAACCCACTCGGGTCCCTTCCTCACTGTGGAAGCTTTGTTCTTTTGCTCTTGC	21724						
QY	669	AAT	AAATCTTTTGCTGCTGCTCACTTCTTTGTGTCAACTACCTTAATGAGCTGTAAACATC	728						
Db	21723	AAT	AAATCTTTTGCTGCTGCTCACTCTTTTGGGTCACTGCTTTAATGAGCTGTAAACATC	21664						
QY	729	ACTG	CGAGGCTGTGGCTTCAATTCCTGAAATCAAC-AGACCAACAAACCCACTGGAAGA	787						
Db	21663	ACG	CAAAAGGTCTGCAAGCTTCACTCTGTAAGCAGACGAAACCAAGAAACCCACCGGAGGA	21604						
QY	788	ACAA	GAATCTCCGATGTCTGCTCTTTAAGACTGTAACTCACTACTGCGAAGCTCTGAG	847						
Db	21603	ACGA	MAATCTCCGATGTCTGCTCTTTAAGACTGTAACTCACTACTGCGAAGCTCTGAG	21545						
QY	848	CTT	CACTCTCGAAGTCAATGAAACCAACCCACAGAAAGGAAGAACTCTGGAACAC	907						
Db	21544	CTT	CACTCTCGAAGTCAATGAAACCAACCCACAGAAAGGAAGAACTCTGGAACAC	21486						
QY	908	CTGA	ATATCTGAAGAAACAACTCCAGACACACTCTTTCAGACTG-TAAACTCTACC	966						
Db	21485	CGAA	ACTTGAAGAAACAACTCCCGAACCCGCTTTAAGAACTGTAACTCACTCAC	21426						
QY	967	GCA	GGGTGTGGCTTCAATTTTGAAGTACAGAACCAACCAACCCGAGAGAAACA	1026						
Db	21425	GCA	GGGTGTGGCTTCAATTTTGAAGTACAGTGAAGCAACCAACCCCAAC	21376						
QY	1027	AAT	TCCAGACAGTAGAAGAAATCTGTAAATTTTGTGTGTGAGTTCCAGGGTAACTCCAGT	1086						
Db	21375	AAT	TCCAGACAGTAGAAGAAATCTGTAAATTTTGTGTGTGAGTTCCAGGGTAACTCCAGT	21348						
QY	1087	CATT	GAAGTCTCAATTGACAGCTTAAGAAACAGAGAAATGTTTGGAGGACACATGTGG	1146						
Db	21347	-----	-----	21348						
QY	1147	GAA	TGTTATGACACAGGCTTGAGATGACATAGGGCAATTTCTGTATCAAACTTAGCTGA	1206						
Db	21347	-----	AGSTTTTGACTTATCTCCATAGG	21324						
QY	1207	AGCA	GGGCCAGGAATATTAATCTAAGAAAGACAGTTTGTGTAGACAGTAGTAGTCTTTGC	1266						
Db	21323	CAC	TGAAAACTTAACAGAAATGATGTAACTTGTATTTTTPAATTAATTAATCAAGTGC	21264						
QY	1267	ATCT	GAGACATGTAGATTAATCAAGCAATTAATTAAGAAAAAATATAGCCAGGTGCATGGC	1326						
Db	21263	ATG	CAACATTAATAGGTTGACTTAAGAAAGCTGGAGA-----GGCTGGCGTGTGTGC	21212						
QY	1327	TCAT	GCTGTAAATCCAGCACTTTGGAGGCCAAAGGGGTGTGTGATCAAGAGTCAAGCGT	1386						
Db	21211	TCAG	GCTGTAAATCCAGCACTTTGGGAGGCTGAGGTGGCGGATCACAAGGTCTAGAGT	21152						
QY	1387	TCGA	GACAGGCTTGACCATATGATGTAACCCCGTCTCTACTTAAAAATTCAAAAATTTAGC	1446						
Db	21151	TCGA	GACAGGCTGTATCAAGTGTAAACCCGTCTGTACTTAAAAACCAAAAAATTTAGC	21092						
QY	1447	CTG	GTGTGTGTGAGCAGCATCTGTAAATCCAG-TACTCAGAGGCTGAGGCAAGGGAATCT	1505						
Db	21091	CGG	GTGTGTGAGCAGCATCTGTAAATCCAGACTCTCAGAGGCTGAGGCA---GATTTG	21035						



```

Oy      1506  CTTGAACCTTGGGACAGAGGTTGCAGTGAAGCCAAAGTATCACACACAGCACTCCATCCG 1565
Db      21034  CTTGAACACGGGAATGAGAGGTTGCATATGAGCCAAAGTTGTGCACATGCACCTCCAGCTTG 20975
Oy      1566  GGTGCACAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAGGAAA 1613
Db      20974  GGCATATAGCGAGATGCATCTTAAAAAAAAAAAAAAAAGAAAAAGA 20927

RESULT 13
US-09-305-384-6
; Sequence 6, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Tyeco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard P
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,649
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-384-6

```

Query Match	20.6%	Score 412.2	DB 3	Length 2834
Best Local Similarity	65.9%	Pred. No. 9.3e-92		
Matches	805	Conservative	0	Mismatches 358; Indels 58; Gaps 12;
Qy	431	TCAGTGGGGA	CTTGAGAGACTTTTGTGTC-----TAGCTAAAGATTTGTAATGC	481
Dp	1370	TCTGTGGGGG	CTTGAGAGATGTTTGTGTGCACTCTGTATCTAGTTAATCTAGTGGGG	1429
Qy	482	ACCAATCAGCA	CTCTGTGTCTTAGCTAAAGATTTGTAATGCAACCAATCAGCACTCTGTAA	541
Dp	1430	ACGTGGAGAA	CTTTGTGTCTTAGCTCAGAGGATTTGTAACGCAACCAATCAGGCCCTGTCA	1489
Qy	542	AA-----	-----TGACCAATTCAGCAGATGTGGGCGGGGTCAATAATGAGAGTAA	587
Dp	1490	AAACAGACCA	CTCGGCTCTACCAATTCAGCAGATGTGGGCGGGGTCAATAATGAGAGTAA	1549
Qy	588	AAACTGGCCAC	CCGAGCCAGCAGAGTGCGCAACCACTCGGGTCCCTTCACACTGTGAAAG	647
Dp	1550	AAGCAGGGTGC	CCGAGCCAGCAGAGTGCGCAACCACTCGGGTCCCTTCACACTGTGAAAG	1609
Qy	648	CTTTGTCTTTT	TGCTCTTACAAATTAATCTTGTCTGTCTCATTTCTTTGTGTCCACACTA	707
Dp	1610	CTTTGTCTTTT	TGCTGTGTGGAATAAATCTTGCTACCTGCTCTTTTGGGTCCACACTG	1669
Qy	708	CCTTTATGAG	CTGTAACTCACTACCTGCGAGGGTCTGTGGCTTCATTTCCGAAGTCAAC-AG	766
Dp	1670	CTTTTATGAG	CTGTAACTCACTACCAACGAGGCTGTGCACTTCCTCTCGAAGCCACTAAG	1729
Qy	767	ACCAAGAAC	CCCACTGTGAAGAACAAAGAACTCCGATGTGTGCTGCTTTAAGAGCTGTAAAC	826
Dp	1730	ACCAAGAAC	CCCACTGTGAAGAACAAAGAACTCCGATGTGTGCTGCTTTAAGAGCTGTAAAC	1788
Qy	827	ACTCAGTGG	AGAGCTGTGACCTTCATCTCCCTGAAGTCAATGAGACCAACCAACCAACA	886
Dp	1789	ACTCAGTGG	AGAGCTGTGACCTTCATCTCCCTGAAGTCAATGAGAGACCAACCAACCAACA	1847
Qy	887	AGGAAGAA	ACTCTGTGACCACTGTGAATATCTGAAGAGAACAACTCCAGACCAACCATCTT	946
Dp	1848	AGGAAGAA	ACTCTGTGACCACTGTGAATATCTGAAGAGAACAACTCCAGAGTACCA-CT	1906
Qy	947	TCAAGACTGT	AACACTCACCGCAAGAGCTGTGTGCTTCATTCTTGAAGTCAAGAACCA	1006

Db	1907	TAAAGCTGTAACTCACTCGAGAGGTCGCGGCTTCTCTTGGAACTCAGTGAACCA	1966
QY	1007	AGAACCCACCGAAGAAACAATTCACAGACAGTAAAGAAATCTGTATTTTGTGTCGTG	1066
Db	1967	AGCACTCACAGTTTCGAGACAAGAGCCAGGAGTTTGATCAGCTCGGCAACATGATG	2026
QY	1067	GCTTCAGAGGTACTCCAGTCATTTGAATCTCCATTCGAGCCTTAAGAAACAGAAATG	1126
Db	2027	AAATCCCTCTCTCGCAAAAAAAAATAATTAACAAAATTTGCGGAGCATGTGTCGCT	2086
QY	1127	GTTTGAAGAGCACATGTGGGAATTTGTATGAACAGGCTTGAGATGCATTAAGGCAAT	1186
Db	2087	GCTGTGCTCCAGTACGCGGAGGCTTAAAGTGGAGAGATCGCTTGAGCCT-GGAGST	2145
QY	1187	TCTGATCAACCTAGCTGAAGCAGGGCCAGGAAATATATCTAAGAGACAGTTTTGTG	1246
Db	2146	GAAAGCTCAGTAGAGCTGTGATTGTACACAGCCCTCTAGGCTGGGGGACAGACT----	2200
QY	1247	TAGACAGTAGTAGTCTTTTGCACTCTGAGACATGTATGATTATCAACAAATTATTAAGAAAA	1306
Db	2201	----GAGACCTCTGTTCCTCCCGCAAAAAAATTTGACAAAGTGTAAATMAAGGTGCCTG	2256
QY	1307	ATATAGCAGGTGCATGCTCATGACCTGTATCCACAGCACTTTGGAGGCCAAGGGTGTG	1366
Db	2257	ATATGGCTAAGCGCAGTGGCTCATGCTGTATTCACAGCACTTTGGAGGCCAAGGGCGGG	2316
QY	1367	TGGATCAC--GAGGTCAAGCGTTGGAACACAGCCTGGCCAAATGTGTAAACCCCGTCTC	1424
Db	2317	CGGGTCACCTAAGGTCAGAGGTGTGAGACGAGCCTGGCCAAATGTGTAAAGCCCATCTC	2376
QY	1425	TACTTAAAAATACAAAAATTAGCCT-----GGTGTGTGTGGCAGCATCTGTATTCCTCA	1476
Db	2377	TTCTTAAAAATACAAAAATTAGCGCGGTGTGGGGGAGGTGTGTGAGCATGCTCTGTATTCCTCA	2436
QY	1477	G-TACTCAGAGGCTGAGGACAGGGGAATCTTTGAACCTTGGAGGACAGAGTTGCAGTGA	1535
Db	2437	GCTACTCAGGAGGCTGAGGACAGGAAATCACTTGAAACCCAGGAGGCGGCGGTTGCAGTGA	2496
QY	1536	GCCAAGATCACACCAAGCACTGCATCC-----TGGGTGACAGAGCGAGACTCTG	1585
Db	2497	GCCGAGATCTGTGCATTGTCACTCCACCACTCCAGCTGGGCAACAAGAGCCAATCTGTG	2556
QY	1586	TCTCAAAAAAATTTTTTTTTTTT 1606	
Db	2557	TCTTAAAAAATTTTTTTTTTTT 2577	

```

RESULT 14
US-09-305-384-5
; Sequence 5, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Selden, Richard F.
; APPLICANT: Heartlein, Michael W.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,649
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 6235
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-384-5

Query Match      20.6%; Score 412.2; DB 3; Length 6235;
Blast Local Similarity 65.9%; Pred. No. 1.3e-98;
Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

```

431 TCAGTGGGAACTTGGAGAACTTTGTGTC-----TAGCTAAGAGTTGTAATGC 481  
482 ACCAATCAGCACTGTGTCTAGCTAAGAGTTGTAATGCAATCAGCACTGTGTA 541  
3319 ACGTGAGAACTTTGTGTAGCTCAGGGATTGTAAAGCACTAAGCGCCCTGTCA 3378  
542 AA-----TGACCAATCAGCAGATGTGGCGGGGTCAATAAAGGAGTAA 587  
3379 AAACAGACCACTGGCTTACCAATCAGAGATGTGGGTGGGCAATTAAGATTA 3438  
588 AAACCTGGCCACCCGAGCAGAGTGGCAACCACTGGGTCCTTCCACACTGTGAAG 647  
3439 AAGCAGGCTGCCGAGCCAGCACTGGCAACGCCAGAGTCCCTATCCCAATATGGCAG 3498  
648 CTTTGTCTTTGCTCTTCACAATTAATCTTGTGCTGTCTCATTTCTTTGTCCACTA 707  
3499 CTTTGTCTTTGCTGTGCTGATTAATCTTGTCTGTCTGTCTTTTGGGTCCACACTG 3558  
708 CTTTATGAGCTTAACTCACTGAGAGGCTGTGGCTTCAATCCTGAAAGTCAAC-AG 766  
3559 CTTTATGAGCTTAACTCACTGAGAGGCTGTGGCTTCACTCTGAAAGCACTAAG 3618  
767 ACCACGAACCCACTGGAAGGAACAAGAACTCCGATGTGCTGCTTTAAGACTGTAA 826  
3619 ACCACGAGCCACCGGAGGAATGAACAACCTCCGCGCGCTGCG-TTAAGACTTAAC 3677  
827 ACTCAGCTGCAAGCTCTGAGCTTCACTCTGAAATGATGAGACCAACCCACAGA 886  
3678 ACTCAGCGCAAGGTCTGAGCTTCACTCT-CAAGCAGAGAGCCAGAACCCACAGA 3736  
887 AGGAAGAACTCTGAGACACTGAAATATCTGAAGGAACAATCCCAACCACTT 946  
3737 AGGAAGAACTCTGAGACACTGAAATATCTGAAGGAACAATCCCAAGTCAACA-CTT 3795  
947 TCAGAGCTTAACTCACTCAGCGCAAGGCTGTGGCTTCACTTCTTGAAGTCAAGAACCA 1006  
3796 TAAGAGCTTAACTCACTCAGCGAGGCTCCGCGCTTCTTCTTGAAGTCAAGAACCA 3855  
1007 AGAACCACCGGAAGGAACAATTCAGACACAGTGAAGAAATCTGATTTTGTCTGTG 1066  
3856 AGACCTCACCAGTTTCGAGACCAAGCCAGAGTTTGAATCAGCTGGGCAACATGATG 3915  
1067 GCTTCCAGGCTTCACTCAGTCAATGAAATCTCCATTGCAAGCTTAAAGAAACAGAGATG 1126  
3916 AAATGCCCTCTCTGCAAAAAAATAAATAAATAATGGCGAGCATGTGTCTCT 3975  
1127 GTTTGAGAGACATGTGGGAATTTTATGGAACAGGCTTGAATGCACTAGAGCATT 1186  
3976 GCTGTGTGTCCCACTACGCGGAGGCTAAAGTGGAGAGATGCTTGAAGCT-GGAGGT 4034  
1187 TCTGATCAAACTAGCTGGAAGCAGGCGCAGAAATTAATCTAAGGAAGACAGTTTGTG 1246  
4035 GAAGACTGAGTGAAGTGTATTTGTAACAAGCCCTCTAGGCTGGGGAGACACT----- 4089  
1247 TAGACAGTGAATGCTTGTGATCTGAGACATGTGATTAATACCAATTAATTGAAGAAA 1306  
4090 -----GAGACCTGTGTTCCCTCCGCAAAAAAATTAAGCAAAAGTGAATTAAGGTGCTG 4145  
1307 AATAGCAGAGTGGATGGCTCATGCTGTAACTCCAGCACTTTGGGAGGCCAAGGGGTG 1366  
4146 AATAGCTAGGCGAGTGGCTCATGCTGTAACTCCAGCACTTTGGGAGGCCAAGGGGTG 4205  
1367 TGGATCAAC--GAGTCAAGCGCTTGAAGCAAGCTGTGCAACATGTGTAAGAAACCCGCTTC 1424  
4206 CGGCTCACTTAAGTCAAGAGTGTGAAGCAAGCTGTGCAACATGTGTAAGAAACCCATCTC 4265  
1425 TACTAAATAATCAAAATTAAGCT-----GGTGTGTGGCAGCATCTGTAACTCCA 1476  
4266 TTTTAAAAAATCAAAATTAAGCTGTGTGGGAGGAGTGTGTGAGAGATGCTGTAACTCCA 4325

1477 G-TACTCAGAGAGCTGAGGAGGGAATCTTTGAACCTTTGGAGGCAAGGTTGCAGTGA 1535  
4326 GCTACTCAGAGAGCTGAGGAGGAGGAATCACTTGAACCCAGAGGCGGGTTGCACTGA 4385  
1536 GCCAATCAGCACTGTGTCTAGCTAAGAGTTGTAATGCAATCAGCACTGTGTA 1585  
4386 GCCGAGTGTGCTCACTGAGCTTCACTCAGCTCAGCTGAGGCAACAGAGCCAACTCTG 4445  
1586 TCTCAAAAAAATAAATAAATAA 1606  
4446 TCTTAAAAAATAAATAAATAA 4466

RESULT 15  
US-09-525-160B-6  
; Sequence 6, Application US/09525160B  
; Patent No. 6569681  
; GENERAL INFORMATION:  
; APPLICANT: Ivanov, Evgenii  
; TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION  
; FILE REFERENCE: 10278/016001  
; CURRENT APPLICATION NUMBER: US/09/525,160B  
; CURRENT FILING DATE: 2000-03-14  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 6235  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-525-160B-6

Query Match 20.6%; Score 412.2; DB 3; Length 6235;  
Best Local Similarity 65.9%; Pred. No. 1.3e-98;  
Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

431 TCAGTGGGAACTTGGAGAACTTTGTGTC-----TAGCTAAGAGTTGTAATGC 481  
3259 TCTGAGTGGGCTTGGAGAAATGTTGTGTCACACTGTATCTAATTGTAATGAGGGG 3318  
482 ACCAATCAGCACTGTGTCTAGCTAAGAGTTGTAATGCAATCAGCACTGTGTA 541  
3319 ACGTGAGAACTTTGTGTAGCTCAGGATTTGTAAGCACTAAGCGCCCTGTCA 3378  
542 AA-----TGACCAATCAGCAGATGTGGCGGGGTCAATAAAGGAGTAA 587  
3379 AAACAGACCACTGGCTTACCAATCAGAGATGTGGGTGGGCAATTAAGATTA 3438  
588 AAACCTGGCCACCCGAGCAGAGTGGCAACCACTGGGTCCTTCCACACTGTGAAG 647  
3439 AAGCAGGCTGCCGAGCCAGCACTGGCAACGCCAGAGTCCCTATCCCAATATGGCAG 3498  
648 CTTTGTCTTTGCTCTTCACAATTAATCTTGTGCTGTCTCATTTCTTTGTCCACTA 707  
3499 CTTTGTCTTTGCTGTGCTGATTAATCTTGTCTGTCTGTCTTTTGGGTCCACACTG 3558  
708 CTTTATGAGCTTAACTCACTGAGAGGCTGTGGCTTCAATCCTGAAAGTCAAC-AG 766  
3559 CTTTATGAGCTTAACTCACTGAGAGGCTGTGGCTTCACTCTGAAAGCACTAAG 3618  
767 ACCACGAACCCACTGGAAGGAACAAGAACTCCGATGTGCTGCTTTAAGACTGTAA 826  
3619 ACCACGAGCCACCGGAGGAATGAACAACCTCCGCGCGCTGCG-TTAAGACTTAAC 3677  
827 ACTCAGCTGCAAGCTCTGAGCTTCACTCTGAAATGATGAGACCAACCCACAGA 886  
3678 ACTCAGCGCAAGGTCTGAGCTTCACTCT-CAAGCAGAGAGCCAGAACCCACAGA 3736  
887 AGGAAGAACTCTGAGACACTGAAATATCTGAAGGAACAATCCCAACCACTT 946  
3737 AGGAAGAACTCTGAGACACTGAAATATCTGAAGGAACAATCCCAAGTCAACA-CTT 3795  
947 TCAGAGCTTAACTCACTCAGCGCAAGGCTGTGGCTTCACTTCTTGAAGTCAAGAACCA 1006

Db 3796 TAAAGCTGTAAACATCTGACGAGGCTCGCGGCTTCTCTTGAAGTCAGTGAGACCA 3855  
QY 1007 AGAACCCAGCGGAAGAAATTCAGACACAGTAGAATCTGTATTTTGTATCTGTG 1066  
Db 3856 AGCATCAACCAATTTCGGAACCAAGCCAGATTGTAGATCAGCTGGGCAATATG 3915  
QY 1067 GCTTCCAGGGTTACTCCAGTCATTTGAAGTCTCCATTGAGCTTTAAGAAACAGAAATG 1126  
Db 3916 AATAGCCCTCTCTGCAAAAAAATTAACAAAAATGGCGGACATGATGTCCTG 3975  
QY 1127 GTTTGAGAGACATCTGGGAATTTTATGACACAGGCTTGAGATCAATAGGCAAT 1186  
Db 3976 GCTGTGTGTCACAGTACGCGGAGGCTAAAGTGGAGATCGCTTAAGCT--GGAGGT 4034  
QY 1187 TGTGATCAACCTAGTGTGAAGACGAGGCAAAATTAATCTTAAGAGACAGTTTGTG 1246  
Db 4035 GAAGATCTGACGTAGCTGTATTTTACCAAGCCCTTGAAGCTGGGGACAGACT----- 4089  
QY 1247 TAGACAGTAGTAGTCTTTGATCTGACATGTAGATTATCAAGCAATTAATTAGAAAA 1306  
Db 4090 ----GAGACCTGTTTCCCTCCGCAAAAAATTTGACAAAAGTATTAAGGTGCTG 4145  
QY 1307 ATATAGCAGGTGATGCTCATGCTGTATTCAGCACTTTGGAGGCAAGGGGTG 1366  
Db 4146 ATATGCTAGGCGCAGATGCTCATGCTGTATTCAGCACTTTGGAGGCAAGGGGTG 4205  
QY 1367 TGTATCAC--GAGGTACAGGCTTCGAGACCAAGCTGGGCAACATGTGAAACCCGCTG 1424  
Db 4206 CGGTCACCTTAAGGTACGAGGTGTGAGACCAAGCTGGGCAACATGTGAAAGCCATCTC 4265  
QY 1425 TACTAAAAATACAAAAATTAAGCT-----GGTGTGTGACAGCATGTATTCCTCA 1476  
Db 4266 TTTCTAAAAATACAAAAATTAAGCTGGGTCAGTGTGAGACATGCTGTATTCCTCA 4325  
QY 1477 G-TACTCAGAGGCTGAGGAGGAAATCTTTGAATTTGAGAGGAGAGGTTTGCATGA 1535  
Db 4326 GCTACTCAGAGGCTGAGGAGGAAATCTTTGAATTTGAGAGGAGGCGGTTGCAGTGA 4385  
QY 1536 GCCAATATCACACACAGCATCTCATTC-----TGGTGTACAGAGCAGACCTGTG 1585  
Db 4386 GCCGAGTCTGTGCTATTCACCTCCACCACTCCAGCTGGGCAACAGAGCCAACTCTG 4445  
QY 1586 TCTCAAAAAAATTAAGCT-----TCTCAAAAAAATTAAGCT-----TCTCAAAAAAATTAAGCT----- 1606  
Db 4446 TCTTAAAAAATTAAGCT-----TCTTAAAAAATTAAGCT-----TCTTAAAAAATTAAGCT----- 4466

RESULT 16  
US-09-305-384-1  
; Sequence 1, Application US/09305384  
; Patent No. 6242218  
; GENERAL INFORMATION:  
; APPLICANT: Treco, Douglas A.  
; APPLICANT: Heartlein, Michael W.  
; APPLICANT: Seiden, Richard F.  
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY  
; FILE REFERENCE: 07236/017001  
; CURRENT APPLICATION NUMBER: US/09/305,384  
; EARLIER FILING DATE: 1999-05-05  
; EARLIER APPLICATION NUMBER: US 60/084,649  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-305-384-1

Query Match 20.6%; Score 412.2; DB 3; Length 6679;  
Best Local Similarity 65.9%; Pred. No. 1,4e-98;  
Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

QY 431 TCAGTGGGAGCTTTGAGAACTTTGTGTC-----TAGCTAAAGATTTGAAATGC 481  
Db 3278 TCTGTGTGGGCTTTGAGAAATGTTTGTGTGACACTGTATCTATGTAATCTAGTGGG 3337  
QY 482 ACCAATCAGACCTCTGTGTAGCTAAAGATTTGAATGACCAATCAGACTCTGTAA 541  
Db 3338 ACCTGAGAACTTTGTGTAGCTCAGGATTTGAACGACCAATCAGCCCTGTCA 3397  
QY 542 AA-----TGACCAATCAGAGATTTGGGGGTCAAAATGAGATTA 587  
Db 3398 AAACAGACCTCGGCTTACCAATCAGAGATTTGGGGGTCAAAATGAGATTA 3457  
QY 588 AAATGGCCACCCGAGCAGACAGTGGCAACCACTGGGCTCCCTTCCACATGTGAG 647  
Db 3458 AACGAGGTGCTCCGAGCCAGCATGCGCAAGCCACAGTCTCTATTCACAAATATGCG 3517  
QY 648 CTTTGTCTTTTGTCTCTTCAAAATTAATCTTGTCTGTCTCATCTTTGTGTCACAT 707  
Db 3518 CTTTGTCTTTTGTCTGTGTCGATTAATCTTGTCTGTCTGTCTTTTGGGTCCACACTG 3577  
QY 708 CTTTATGAGCTGTAACTCATCTGACAGGATTTGTGCTTCAATCTGAAATCAAC-AG 766  
Db 3578 CTTTATGAGCTGTAACTCATCTGACAGGATTTGTGCTTCAATCTGAAATCAAC-AG 3637  
QY 767 ACCAGCAACCCACTGGAAGGAACAAAGAACTCCGATGTGCTGCTTTAAGCTGTAC 826  
Db 3638 ACCAGCAACCCACTGGAAGGAACAAAGAACTCCGATGTGCTGCTTTAAGCTGTAC 3696  
QY 827 ACTCACTGCGAAGCTGTGACCTTCACTCTGTAAGTCAAGTGAACCAAAACCCACAGA 886  
Db 3697 ACTCACTGCGAAGCTGTGACCTTCACTCTGTAAGTCAAGTGAACCAAAACCCACAGA 3755  
QY 887 AGAAGAACTCTGACACACCTGAAATTTGAAAGAACAACTCAGACACCACTTT 946  
Db 3756 AGAAGAACTCTGACACACCTGAAATTTGAAAGAACAACTCAGACACCACTTT 3814  
QY 947 TCAGCTGTAACTCACCAGGAGGTGTGAGCTTCAATCTTGAAGTGAAGACCA 1006  
Db 3815 TTAGACTGTAACTCACCAGGAGGTGTGAGCTTCAATCTTGAAGTGAAGACCA 3874  
QY 1007 AGAACCCAGCGGAAGAAATTAATTCAGACAGTGAATTTGATGTCGTG 1066  
Db 3875 AGCATCAACAGTTTGTGACACAAAGCCAGAGTTTGAATGACCTGGGCAATGATG 3934  
QY 1067 GCTTCCAGGGTTACTCCAGTCATTTGAATCTCATTTGAGACCTTTAAGAAACAGAAATG 1126  
Db 3935 AATAGCCCTCTCTGCAAAAAAATTAACAAAAATTTGCGGAGCATGTGTCTCGT 3994  
QY 1127 GTTTGAGAGGACATGTGGGAATTTGATGACAGGCTTGAAGTGAATGAGGCAAT 1186  
Db 3995 GCTGTGTGTCAGCTACGCGGAGGCTTAAAGTGGAGATGCTTGAAGCT--GGAGGT 4053  
QY 1187 TGTGATCAACCTAGCTGGAAGGAGGCGAGAAATTAATCTTAAGAAAGACAGTTTGTG 1246  
Db 4054 GAAGATCTGACGTAGCTGTGATTTGACCAAGCCCTCTAGGCTGGGGACAGACT----- 4108  
QY 1247 TGAACAGTAGTAGTCTTTGATGTGAATGTATTAAGCAATTAATTAGAAAAA 1306  
Db 4109 ----GAGACCTGTTTCCCTCCGCAAAAAATTTGCAAAAGTATTAAGAGTCCCTG 4164  
QY 1307 ATATAGCAGGTGAGTGTGCTATGCTGTATATCCAGACCTTTGGAGGCAAGGGGTG 1366  
Db 4165 ATATGCTAGGCGCAGATGCTCATGCTGTATATCCAGACCTTTGGAGGCGAGGCGG 4224  
QY 1367 TGTATCAC--GAGGTACAGGCTTTCAGACCAAGCTTGGCCAACTGTGTGAAACCCGCTG 1424  
Db 4225 CGGTCACCTTAAGGTCAAGAGTGTGAGCAAGCTTGGCCAACTGTGTGAAACCCGCTG 4284  
QY 1425 TACTAAAAATACAAAAATTAAGCT-----GATGTGTGACAGCATCTGTATTCCTCA 1476  
Db 4285 TTTCTAAAAATACAAAAATTAAGCCGCTGTGGGGCAATGTGTGAGCATGCTGTATTCCTCA 4344  
QY 1477 G-TACTCAGAGGCTGAGGAGGAGAACTCTTGAATTTGAGAGGAGAGGTTGCAGTGA 1535

DB 4345 GCTACTCAGGAGGCTGAGGCGAGAGATCACTTGAACCCAGAGGGCGGCTTGACAGTGA 4404  
1536 GCCAAGATCAACACACGACATCTCCATCC-----TGGGTGACAGAGCCAGACTCTG 1585  
DB 4405 GCCGAGATGCTGCACTTGACCTCCACCTCCAGCTGGGCAACAAAGCCAAACTCTG 4464  
QY 1586 TCTCAAAAAAAAAAAAAAAAAA 1606  
DB 4465 TCTTAAAAAAAAAAAAAAAA 4485

RESULT 17  
US-09-525-160B-5  
; Sequence 5, Application US/09525160B  
; Patent No. 6569681  
; GENERAL INFORMATION:  
; APPLICANT: Ivanov, Evgenii  
; TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION  
; FILE REFERENCE: 10278/016001  
; CURRENT APPLICATION NUMBER: US/09/525,160B  
; PRIORITY FILING DATE: 2000-03-14  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 6679  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-525-160B-5

Query Match 20.6%; Score 412.2; DB 3; Length 6679;  
Best Local Similarity 65.9%; Pred. No. 1,4e-98;

Matches 805; Conservative 0; Mismatches 358; Indels 58; Gaps 12;

QY 431 TCAGTGGGAGCTTGGAGAACTTTGTGTC-----TAGCTAAGATTTGAATGC 481  
DB 3278 TCTGGTGGGCGCTTGGAGAACTTTGTGTCGACACTCTGTATCTAGTTATCTAGTGGG 3337  
QY 482 ACCAATCAGACCTCTGTCTAGCTAAGATTTGAATGCACCAATCAGACCTCTGTA 541  
DB 3338 ACGTGGAGAACTTTGTGCTAGCTCAGGATTTGAACGCAACCAATCAGCGCTGTCA 3397  
QY 542 AA-----TGACCAATCAGCAGATGTGGCGGGTCCCAATTAAGAGATTA 587  
DB 3398 AAACAGACCACTCGGCTCTACCAATCAGCAGATGTGGTGGGCGGAGATTAAGATTA 3457  
QY 588 AAATGGCCACCCGAGCGAGAGAGTGGACCCACTCGGGTCCCTTCACACTGTGAG 647  
DB 3458 AAGCAGGCTGCGGAGCGAGAGTGGACCGGCAAGTCCATATCCAAATATGAGAG 3517  
QY 648 CTTTGTCTTTGCTCTTCAATTAATCTTGTCTGTCTCATTTTGTGTCCACTA 707  
DB 3518 CTTTGTCTTTGCTCTTCAATTAATCTTGTCTGTCTCATTTTGTGTCCACTG 3577  
QY 708 CCTTATGAGGTGAACACTCACTGCGAGGCTGTGGCTTCATTCTGAAGTCAAC-AG 766  
DB 3578 CTTTATGAGGTGAACACTCACTCAAGAGTCTGCACTTCACTCTGGAAGCCACTAAG 3637  
QY 767 ACCAAGAACCCACTGGAAGAAAGAAAGAACTCCGATGTGCTGCTTTAAGCTGTAC 826  
DB 3638 ACCAAGAACCCACTGGAAGAAAGAAAGAACTCCGCGCGCTGCC-TTAAAGCTATVAC 3696  
QY 827 ACTCATGCGAAGCTCTGAGCTTCACTCTGAAAGTCACTGAGACCAAAACCAACGA 886  
DB 3697 ACTCATGCGAAGCTCTGAGCTTCACTCTGAAAGTCACTGAGACCAAAACCAACGA 3755  
QY 887 AGGAAGAACTCTGAGACACCTGAATATCTGAAGAAACAACTCCAGACACCATCTT 946  
DB 3756 AGGAAGAACTCTGAGACACCTGAATATCTGAAGAAACAACTCCAGATGACCA-CTT 3814  
QY 947 TCAGAGCTGTAACTCAACCGCAAGGCTGTGCTTCAATCTTGAAGTCAAGACCA 1006  
DB 3815 TAAAGCTGTAACTCACTGAGAGGCTCCGCGGCTTCTTGAAGTCAAGTGAACCA 3874

QY 1007 AGAACCCACCGAAGAAACAAATTCACAGACAGTAGAAATCTGTATTTTGTATCTG 1066  
DB 3875 AGCACTACACAGTTTCGACACAAAGCCAGAGATTTTGATACAGCTGGGCAACATGATG 3934  
QY 1067 GCTTCAGAGGTACTCCAGTCACTGAATCTTCATTTGAAGCTTAAGAAACAGAAATG 1126  
DB 3935 AAATCCCTCTCTGCAAAAAAAAAAAAAAAAAAATTAACAAAATTTGGCGAGATGCTCCGT 3994  
QY 1127 GTTTGAGAGACACATGTGGAAATTTGTATGACACAGCTTGAAGTACATATAGGCAAT 1186  
DB 3995 GCTTGTGCTCCAGTACCGGAGGCTTAAAGTGGAGAGATGCTTTAGAGCT-GGAGAGT 4053  
QY 1187 TCTGATCAAACTTACCTGGAAGCAGAGGCGCAAGAAATATATCTAAGAAAGACATTTTG 1246  
DB 4054 GAAAGCTGAGTAGACTGTGATTTGACCAAGCCCTTAGGCTGGGGGACAGACT----- 4108  
QY 1247 TAGACAGTAGTAGTCTTGTGACATCTGACATCTGATTAATTAAGCAATTAATTAAGAAA 1306  
DB 4109 ----GAGACCTGTGTTCCCTCCGCAAAAAAAAAATTAACAAAGTATTAAGAGTGCTG 4164  
QY 1307 ATATAGCCAGTGCATGCTCATGCTGTATATCCAGACCTTTGGAGGCCAAGGGGTG 1366  
DB 4165 ATATGCTTAGGCGCACTGCTCATGCTGTATATCCAGACTTTGGAAAGCCAGAGCGGG 4224  
QY 1367 TGGATCAC--GAGGTGAGCGTTTGAACACAGCTTGGCCACATGCTGAACCCGCTCTC 1424  
DB 4225 CGGGTCACTTAAGTCAAGAGTGAACACAGCTTGGCCACATGAGAAACCCATCTC 4284  
QY 1425 TACTAATAATCAAAATTAAGCT-----GGTGTGTGCGACGCAATCTGTAATCCCA 1476  
DB 4285 TTCTAATAATCAAAATTAAGCTGGCTGTGGGCGAGTGTGAAGCATGCTGTATATCCCA 4344  
QY 1477 G-TACTCAGAGGCTGAGGCGAGGAACTCTTGAATTTGGAGGCGAGAGTTGCAAGTA 1535  
DB 4345 GCTACTCAGAGGCTGAGGCGAGGAAATCACTTGAACCAAGAGGCGGCTTGCAGTA 4404  
QY 1536 GCCAAGTCAACACACGACATCTCCATCC-----TGGGTGACAGAGCAACTCTG 1585  
DB 4405 GCCGAGATGTGCTATTTGACCTCAACCACTCCAGCTGGGCAACAAAGCCAAACTCTG 4464  
QY 1586 TCTCAAAAAAAAAAAAAAAAAA 1606  
DB 4465 TCTTAAAAAAAAAAAAAAAA 4485

RESULT 18  
US-09-949-016-12298  
; Sequence 12298, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTNER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIORITY FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIORITY FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIORITY FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12298  
; LENGTH: 115963  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(115963)  
; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-12298

Query Match 20.6%; Score 411.8; DB 3; Length 115963;  
Best Local Similarity 68.5%; Pred. No. 6.1e-98;  
Matches 776; Conservative 0; Mismatches 282; Indels 75; Gaps 12;

QY 493 CTCTGTGTCTAGCTAAAGATTGTAAATGCAACCAATCAGACTCTGTAA----- 543  
DB 74942 CTTTGTCTAGCTAGGGATTGTAACTCAACCAATCAGCGCCTTGCAAAACAGACAC 75001

QY 544 -----TGACCAATCGCAGGATGTGGCGGGCTCAATTAAGGAGTAAACCTGCCAC 598  
DB 75002 TGGGCTCTACCAATCAACAGGATGTGGGCGGCGCATTAAGAAATTAAGCAGGCTGC 75061

QY 599 CCGAGCAGAGGAGGGAACCACTCGGGTCCCTTCACACTGTGAAAGTTGTTCTT 658  
DB 75062 CCGAGCAGAGGAGGGAACCGCTCGGGTCTCTTCGCACTGTGAAAGTTGTTCTT 75121

QY 659 TGTCTTTCACAAATTAATCTTGTGCTCATCTTGTGTCCACACTACTTATGAGC 718  
DB 75122 TGC-----TGTTGGTCCCACTGCTTTATGAGC 75152

QY 719 TGTACACTCTACTCGAGGCTGTGGCTTCACTTCTGAATGTAAC-AGACACGAACTC 777  
DB 75153 TGTACACTCTACTCGAGGCTGTGGCTTCACTTCTGAAGACGAGACGAGCC 75212

QY 778 ACTGGAAGGAACAAAGACTCCGATGTGCTGCTTAAGTGTAACTCACTGCGA 837  
DB 75213 ACCAGAGGAACAAAGACTCCAGATGCGCAC-CTTAAGAGTGTAACTCACTGCGA 75271

QY 838 AGCTGTGACCTTCACTCTGAAAGTCACTGAGACCAACACCAAGAGAAAGT 897  
DB 75272 AGCTGTGACCTTCACTCTGAAAGTCACTGAGACCAACACCAAGAGAAAGT 75330

QY 898 CTGGAACAACCTGAATATCTGAAGAAACAACTCCAGACCAACCTTTCAGAGCTGA 957  
DB 75331 CCAACACAACCTGAATATCTGAAGAAACAACTCCAGAGCGCA-CTTTAAGAGCTGA 75389

QY 958 AACTCTACCGCAAGGCTGTGGCTTCACTTGTGAAGTCAAGCAAGCAAGAACCACTG 1017  
DB 75390 AACTCTACCGCAAGGCTGTGGCTTCACTTGTGAAGTCAAGCAAGCAAGAACCACTG 75448

QY 1018 GAAGGAACAAATTCAGACAGTAGAATCTGTATTTTGTATCTGTGCTCCAGGGT 1077  
DB 75449 -----AATTCGACACAGAGGACCAATTAAGGCTCAATTAAGAGAGATTGG 75499

QY 1078 TACTCCAGTCAATGAAGTCTCAATGCACTTAAGGAAACAGAGATGTTTGGAGAG 1137  
DB 75500 TGGCTCAATTAAGAGTGTGTTAGTAGAAGAGTGTGGAGAAATTTCTAGAGTA 75559

QY 1138 CACATGTGGGAATGTATGAGACAGGCTTGAATGACATAGGCAATTTCTGATCAAC 1197  
DB 75560 TTTAGGAATTAAGATGATTCCTGAACATTAAGGGAAGAAATAGGGGCG-AT 75612

QY 1198 CTAGCTGAAGAGGCGCAGAAATTAATTAAGAGACAGTTTGTAGCACTAGT 1257  
DB 75613 AAGGCTGGCAAGTACTCTGAAGAGATTTCTGTATCAAGGAAGTTTGGAAATTTGT 75672

QY 1258 AGCTTTGATCTGAACATGTAGATTAATCAAGCAATTAATTAAGAAATTAAGCCAG 1317  
DB 75673 ATTTGTCTATTTTGA-----AGTCAACAACATATTAAGAGCCCTTAGGGGTGG 75725

QY 1318 TGGCATGTCTATGCTCTGAATCCAGCACTTTGGAGGCGCAAGGGGTGTGATCAG--G 1375  
DB 75726 TATGTGTCTACGCTGTAAATTCACACTTTGGAGGCTGAGGCAAGTGTATCACTG 75785

QY 1376 AGGTGAGGCTTGAAGCAGGCTGGCAACATGTGAACCCGCTCTTAATAAAATA 1435  
DB 75786 AGGTGAGGCTTGAAGCAGGCTGGCAACATGTGAACCTTGGCTCTTAATAAAATA 75845

QY 1436 C-AAAAATTAAGCTGTGTGTGTGACGCACTGTATATCCAG-TACTCAGAGGCTGAG 1493  
DB 75846 CAAAAAATTAAGCTGTGTGTGTGACGCACTGTATATCTAAGCTACTCAGAGGCTGAG 75905

QY 1494 GCAGGGAATCTCTTAACCTTGAGGAGCAGAGGTTGACATGAGCCAAAGATCACACCAG 1553  
DB 75906 GTAGGAATCTCTTAACCTTGAGGAGCAGAGGTTGACATGAGCCAAAGATCTGCAATTG 75965

QY 1554 CACTCCATCTGGGTGACAGAGGAGCTGTCTCAAAAAAAAAAAAAAAAAAAAA 1606  
DB 75966 CACTCCAGCTGGGCGACAGAGTGAAGCTGTCTCAAAAAAAAAAAAAAAAAAAAA 76018

RESULT 19  
US-09-949-016-12707/c  
; Sequence 12707, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12707  
; LENGTH: 190078  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc. feature  
; LOCATION: (1) ... (190078)  
; OTHER INFORMATION: n = A, T, C or G  
US-09-949-016-12707

Query Match 20.3%; Score 407.2; DB 3; Length 190078;  
Best Local Similarity 84.2%; Pred. No. 1.2e-96;  
Matches 497; Conservative 0; Mismatches 83; Indels 10; Gaps 3;

QY 437 GGGGACTTGGAACTTTTGTCTAGCTTAAGATGTAAATGCAACCAATCAGACTCT 496  
DB 57447 GGGGACTTGGAACTTTTGTCTAGCTTAAGATGTAAATGCAACCAATCAGACTCT 57388

QY 497 GT-----GTAGCTAAAGATTTGTAAATGACCAATCAGACTCTGTAAATGAC 548  
DB 57387 GTCAAAACGAGCACTCAGACTCTGTAAATGACCAATCAGACTCTGTAAATGAC 57328

QY 549 CAATGACAGATGTGGGCGGGGTCAATTAAGAGATTAAGAGGCAATGCGCAC- CGAGCCAG 607  
DB 57327 CAGTCAAGATTAATGTGGGTGGGCAATTAAGAGATTAAGAGGCAATGCGCAC- CGAGCCAG 57268

QY 608 CAGTGGCAACCACTCGGGTCCCCCTTCCACATGTGGAAGCTTTTGTCTTTGCTTGA 667  
DB 57267 CTGCGCAACCGCTCTGTGTCTCTTCCACGAGTGTGAAGCTTTTGTCTTTGCTTGA 57208

QY 668 CAATAATCTTCTGCTGCTCATCTTTTGTGTCACACATCACTTATAGAGCTGTAACT 727  
DB 57207 CAATAATCTTCTGCTGCTCATCTTTTGTGTCACACATCACTTATAGAGCTGTAACT 57148

QY 728 CACTCGAGAGGCTGTGTGCTTCAATCTGAAGTCAAC-AGACCAGAACCACTGGAAG 786  
DB 57147 CCTCGAAAGGCTGTGAGCTTCACTTGAAGGCGCAGCAGACCAAAACCAAGAGG 57088

QY 787 AACAAAGATCCGATGTGCTCTTAAGAGCTGTAACTCACTGGAAGCTGTGA 846  
DB 57087 AACAAAGATCCGATGTGCTCTTAAGAGCTGTAACTCACTGGAAGCTGTGA 57028

QY 847 GCTTCACTCTGAAGTCAAGTGAACCAAAACCAAGAGAAAGAACTGTGACACA 906  
DB 57028 GCTTCACTCTGAAGTCAAGTGAACCAAAACCAAGAGAAAGAACTGTGACACA 906

Accession	Sequence	Position
Dd	57027 GCTTCATTTCCGGAAGTCAGGAGACCCAGAACCCACGAGAGAAAGAACTCTGACACA	56968
Oy	907 CCGAATATCTGAGAGAAACAACTCCGAGACACGATCTTTCAGACTGTACACATCCACC	966
Dd	56967 TCTAATACCTCTGAGAGAACCAACTCTGACACACACATCTTTAAGAACTATACATCACT	56908
Oy	967 GCAAGGGCTCTGCGCTTCATTTCTTGAAGTCGACAGACCAAGAACCCACC	1016
Dd	56907 GCGAGGGCTCTGCGCTTCATTTCTTGAATCGCGCTAGCCAGAAAGCCACC	56858

RESULT 20  
US-09-949-016-17026/c

```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17026
; LENGTH: 190078
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(190078)
; OTHER INFORMATION: n = A,T,C OR G
;
US-09-949-016-17026

```

Query Match	20.3%	Score 407.2	DB 3	Length 190078
Best Local Similarity	84.2%	Pred. 1.2e-96		
Matches 497	Conservative	0	Mismatches 83	Indels 10
			Gapd	3
Qy	437	GGGAGCTTGAGAACTTTGTGCTAGCTAAAGATTGTAATGCACCAATCAGACTT	496	
Db	57447	GGGAGCTTGAGAACTTTGTGCTAGCTAAAGTTTGAACGCAATCAGACTCT	57388	
Qy	497	GT-----GTACGTAAAGATTGTAATGCACCAATCAGACTCTGTAATAATGAC	548	
Db	57387	GTCAAAACGACCACTAGCACTCTGTAATAATGACCAATCAGACTCTGTAATAATGAC	57328	
Qy	549	CAATCAGCAGGATGTGGCGGGGTCAATAAAGGAGTAATAAATTGGCAAC-CGAGCCAG	607	
Db	57327	CAGTCAGATTAATGTGGGTGGGGCCAAATAAAGGATAAAGGCAGGCCACGGAGCCAG	57268	
Qy	608	CAGTGGCAACCACTCGGGTCCCTTCCACACTGTGGAAAGCTTTGTCTTTTGCTTCA	667	
Db	57267	CTGGGGCAACCCCGCTCTGGTCTCCCTTCCACCGGTGGAAAGCTTTGTCTTTTGCTTCA	57208	
Qy	668	CAATAAATCTTGCTGCTCTCAATTTTGTGTCACACACTTATATAGCTGTAACT	727	
Db	57207	CAATAAATCTTGCTGCTCTCACTTTTGGTCCGGTCACTTTATAGAGCTGTAACT	57148	
Qy	728	CATGCGAGGCTCTGTGCTTCATTCCTGAAATCAAC-AGACCAAGAACCCACTGGAAG	786	
Db	57147	CCCTGCAAAAGGTCTGCACCTTCATCTTGAGGCGACGCGAGCAACAAACCAAGAAAG	57088	
Qy	787	AACAAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCACTGCGAAGCTTGA	846	
Db	57087	AACGAAACAATCAGACACGCTGCTTTAAGAGCTGTAACTCACTCAACCGTGAAGTTCGA	57028	
Qy	847	GCTTCATCTCTGAAAGTCAAGTAGAACCAACCAACCAAGAGAAATCTTGACACA	906	

Accession	Sequence	Position
Db	57027 GCCTTCATTCCTGAAGTCAGGAGACCAACGACCACCAAGAAAGAAACTCTGGACACA	56968
Qy	907 CCTGATATATCTGAAGAACAAATCTCCAGACACACCATCTTTGAGAGCTTAAACATCCACC	966
Db	56967 TCTGAACATCTGAAGAACAAATCTGGACACACCATCTTTAAGAACTTAAACATCTACT	56908
Qy	967 GCAAGGCTCTGTGGCTTCATCTCTGAAGTCAGCAAGACCAAGAACCCACC	1016
Db	56907 GCGAGGGTGTGTGCTTCATCTCTGAAGATGCGCTAAGACCAAGAACCCACC	56858

RESULT 21

```

/ Sequence 184449; Application US/02237020
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 184449
/
/ LENGTH: 601
/
/ TYPE: DNA
/
/ ORGANISM: Human
/
/ US-09-949-016-184449

```

Query Match	20.3%;	Score 406.8;	DB 3;	length 601;
Best Local Similarity	84.1%;	Pred. No. 1.3e-97;		
Matches 496;	Conservative 1;	Mismatches 83;	Indels 10;	Gaps 3;

Qy	437	GGGAGCTTGGAGAACCTTTGTGTCTAGCTAAAGAAATGGAATGACCAATCAGACCTC	496
Db	7	GGGAGCTTGAAGAACTTTGTGTCTAGCTAAAGAAATTTGTAACCGAACCAATCAGACCTC	66
Qy	497	GT-----GTCTAGCTAAAGAAATGTAATGCAACCATCAGCACTCTGTAAAAATGAC	548
Db	67	GTCAAAAGGACCACTCAGCACTCTGTAAAAATGACCCGACTCCTCTGTAAAAATGAC	126
Qy	549	CAATCAGCAGATGTGGGCGGGGTCAAAATAAGGAGTAATAACTGGCCACC--CGAGCCAG	607
Db	127	CACCTCAGCAATTATGTGTGGTGGGGCCAAATAAAGGAATTAAGGAGGCCAACCGGAGCCAG	186
Qy	608	CAGTGGCAACCCACTGGGTCCCTCCCTCCACACTGTGGAAGCTTGTTCCTTTCCTTCA	667
Db	187	CTGGCGCAACCCGCTCTGGTCTCCTCCAGAGGTGTGAAGCTTGTTCCTTTCGCTCTTCA	246
Qy	668	CAATTAATCTTGCTGCTCATTTCTTGTGTGCACACTACTTTATGAGCTGTAACT	727
Db	247	CAATTAATCTTGCTGCTCATCTCTTTGTGGTCCGGTCACTCTTATGAGCTGTAACT	306
Qy	728	CACGTGAGGGGTCTGTGGCTTCACTTCGTGAAGTCAAC-AGACCAAGAACCCACTGGAAAG	786
Db	307	CCCTGCAAGAGTCTGACACTTCTCTAGGCGACAGAACCAACAAACCCACAGAAAG	366
Qy	787	AACAAAGAACTCCCGATGTGCTCCCTTTAAGACTGTAACTCTCACTCGAGAGCTTGCA	846
Db	367	AAGAAACAACTCCAGACAGCTCTCTTTAAGAGCTCAACACTCACCGTGAAGTCTGCA	426
Qy	847	GCTTCACTCTGTGAAGTCAGTGAACCAACCAACCCACAGAAGGAAGAAACTCTGGAACA	906
Db	427	GCTTCATTTCTGGAAGTCAAGGAACCAAGAACCCACAGAGGAAGAACTCTGGAACCA	486



QY 907 CTTGAATATCTGAAGAAACAACCTCCAGACACCAATCTTTAGAGCTGTAACTCACC 966  
| | | | |  
DB 487 TCTGAACATCTGAAGAAACAACCTCCAGACACCAATCTTTAAAGTATTAACCTCATC 546  
| | | | |  
QY 967 GCAGAGGCTGTGCTTCACTTTTGAAGTACAGCAAGCCGAAGACCAC 1016  
| | | | |  
DB 547 GCGAGGCTGTGCTTCACTTTTGAAGTACAGCAAGCCGAAGACCAC 596  
| | | | |

## RESULT 22

US-09-949-016-14890/c  
; Sequence 14890, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 14890  
; LENGTH: 93364  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(93364)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14890

Query Match 20.2%; Score 404.8; DB 3; Length 93364;  
Best Local Similarity 67.6%; Pred. No. 3,9e-96;  
Matches 719; Conservative 0; Mismatches 277; Indels 68; Gaps 8;

QY 431 TCAGGTGGGACCTTGGAGAACTTTGTCTAGCTTAAGGATTTAAATGACCAATCAG 490  
| | | | |  
DB 18468 TCTAGTGGGACCTTGGAGAACTTTGTCTAGCTTAAGGATTTAAATGACCAATCAG 18409  
| | | | |  
QY 491 CACTGTGTCTAGCTTAAGGATTTAAATGACCAATCAGCA----- 533  
| | | | |  
DB 18408 CACTGTGTCTAGCTTAAGGATTTAAATGACCAATCAGCA----- 18349  
| | | | |  
QY 534 -----CTCTGTAAATGACCAATCAGCAAGATGTGGCGGGCTCAATTAAGGAGT 585  
| | | | |  
DB 18348 AATCAGCTCTCTGTAAATGAGGCAATCAGCAAGATGTGGCGGGCTCAATTAAGGAGT 18289  
| | | | |  
QY 586 AAAAAATGCGACCCGAGCGAGGAGGCAACCACTCGGGTCCCTTCCACACTGTGA 645  
| | | | |  
DB 18288 AAAAAAGAGCTGCGAGCGAGGAGGCAACCACTCGGGTCCCTTCCACACTGTGA 18229  
| | | | |  
QY 646 AGCTTGTCTTTGCTCTTCACAATTAATCTTGTCTGCTCATTTTGTGTCCAGC 705  
| | | | |  
DB 18228 AGCTTGTCTTTGCTCTTCACAATTAATCTTGTCTGCTCATTTTGTGTGTCCAGC 18169  
| | | | |  
QY 706 TACCTTTATGAGCTGTAACTCACTGCGAGGCTCTGTGCTTCAATTCCTGAAGTCAAC- 764  
| | | | |  
DB 18168 TACCTTTATGAGCTGTAACTCACTGCGAGGCTCTGTGCTTCAATTCCTGAAGTCAAC- 18109  
| | | | |  
QY 765 AGACACGAACCACTGAAGAAACAAGAACTCCGAGTGTGCTG----- 809  
| | | | |  
DB 18108 AGATACGAACCACTGAAGAAACAAGAACTCTGAGCGGAGGAAACAACACTTCCAG 18049  
| | | | |  
QY 810 -----CCTTTAAGAGCTGTAACTCACTGCGAGGCTCTGTGAGCTTCACTCTGAAGT 862  
| | | | |  
DB 18048 AGTGCACACTTAAGAGCTGTAACTCACTGCGAGGCTCTGTGAGCTTCACTCTGAAGC 17989  
| | | | |

QY 863 CAGTGAGACCAACAACCCAGAGAAAGAACTCTGACACACCTGAATATCTGAAG 922  
| | | | |  
DB 17988 CAGTGAGACCAACAACCCAGAGAAAGAACTCTGAACATGTCCGAACATCAGAGG 17929  
| | | | |  
QY 923 AACAACTCCAGACACCACTTTTCAAGCTGTAACTACACCGCAAGGCTGTGGCT 982  
| | | | |  
DB 17928 AAGAACTCTGACACCGCTATCTTTAAGAACTTAACCTACACCGCTAGAGGCTGGGGCT 17869  
| | | | |  
QY 983 TCATTTCTGAAGTCAAGACCAAGAACCCACCGAAGGAACAAATTCAGACACGTA 1042  
| | | | |  
DB 17868 TCATTTCTGAAGTCAAGTGAACCAAGAACCCAC-----AATTCCGACATCTACTA 17819  
| | | | |  
QY 1043 GGAATCTGATTTTGTATCTGTGCTTCAGGGTACTCACTCACTGAATGATCTCAT 1102  
| | | | |  
DB 17818 GCTACTGAGAGGCTGAGTGGAGATC--GCTTGAAGCTGGAGGCGCAAGCTGCAAT 17761  
| | | | |  
QY 1103 GCAGCTTAAGAAACAGAAATGTTTGGAGAGACATGTGGAAATTTTATGACCA 1162  
| | | | |  
DB 17760 GAGTGATGATGATGATGCA--CTACTATCTCCAGCTGGGTGACAGACATGCGCT 17702  
| | | | |  
QY 1163 GCTTGAATGACATAGGCGATTTGTATCAACCTA-----CTGGAAGCAGGGCGAG 1217  
| | | | |  
DB 17701 GCTCCAGGAAAAAATGTTGTGTATGTATGTATGCGGTATGTATCTGTG 17642  
| | | | |  
QY 1218 GAAATATATCTAAGAGACAGTCTTTTGAAGCACTAGTACTTTGATCTGACAT 1277  
| | | | |  
DB 17641 TATATAGTATTAACAAACAGCAGCAATTAAGACAGATCTGTCCACAGACAAAG 17582  
| | | | |  
QY 1278 GTAGATTATCAAGCAATTAATTAAGAAATATAGCAGGTGATGAGTCTGATG 1337  
| | | | |  
DB 17581 AATGAACAGAAAGATTAATAAACAACAGACAGCAGGATGTGTCTATGCTGTGA 17522  
| | | | |  
QY 1338 ATCCAGCACTTTGGAGGCCAAGGGTGTGAT--CAGAGTCAAGCGCTTGAAGCA 1395  
| | | | |  
DB 17521 ATCCAGCACTTTGGAGGGAGTGGAGAGATTCCTTGAAGTCAAGAGTTGAAGCA 17462  
| | | | |  
QY 1396 GCTGGCCAAACATGTGAAACCCGCTCTACTTAAATAATCAAA 1439  
| | | | |  
DB 17461 GCTTGAAGCAATATGAGACACCACTCTTATTTCTGAATTAATA 17418  
| | | | |

## RESULT 23

US-09-949-016-12707  
; Sequence 12707, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 12707  
; LENGTH: 190078  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(190078)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12707

Query Match 20.2%; Score 403.8; DB 3; Length 190078;  
Best Local Similarity 84.8%; Pred. No. 9,8e-96;



OTHER INFORMATION: n = A, T, C or G  
US-09-949-016-17026

Best Local Similarity 84.8%; Pred. No. 9.8e-96;  
Matches 513; Conservative 0; Mismatches 82

447 AGAAGCTTTGTCCTAGCTAAAGATTGTAATGCACCAATCAGACTCTGT-----GT 500

**F01** CTTCTGTTTAAATTCCTCCACCAATTCTGTGAAGAAAATAAAGAATGGACCGTAATCAGCACGGA 560  
/508 AGCACCCCCGAGGCCTAACCAAAGGATGGAGCACAAGCAACCAACCTCATCTCTGACTTGAAAACTTAA 750

561 TGTGGACCGGGTCAATAAAGGAGTAAAACTGGCCACCCGAGCCAGCACTGGCAACCCA 620

621 CTCGGGTCCCTTCCACACTGTGGAGCTTTGTTCTTTGCTCTTCACATAAATCTTGC 680

681 TGCTGTCATTCTTGTGTCACACTACCTTATGAGCTGTAACTCCTGCGGGGTC 740

741 TGTGGCTTCATTCTCGAAGTCAAC-AGACCAGCAACCACCTGGAAGGAACAAGAAGTCC 799

[illegible]

7924 ACTCAGCGAGACCCAGCAACCCACTGGAAGGAAGAACTCCGATACATCTGACATCTGA 7988

7984 AGGAACAACCTCGGACACACCATCTTTAAGACTGTACACTCACCAAGAGGTCCTG 804

8044 GCTTCATTCTAGAGTCAAG-AGACCAAGAACCCACCAAGAGAACCAATTCGGACACA 8100

Patent No. 6943241  
GENERAL INFORMATION

FILE REFERENCE: HI-AU105  
CURRENT APPLICATION NUMBER: US/10/104,047

NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1

```

; TYPE: DNA
; ORGANISM: Homo sapiens

```

Query Match	20.0%	Score 401.2;	DB 3;	Length 2942;
-------------	-------	--------------	-------	--------------

Query Match	20.0%;	Score 401.2;	DB 3;	Length 2942;
-------------	--------	--------------	-------	--------------



Query Match 19.1%; Score 382; DB 3; Length 80355;  
Best Local Similarity 66.4%; Pred. No. 3.9e-90;  
Matches 785; Conservative 0; Mismatches 210; Indels 187; Gaps 9;

437 GGGGACTTGGAGAACTTTTGTCTAGCTAAAGATTGTAATGACCAATCAGACCTCT 496  
|||||  
8672 GGGGACTTGGAGAACTTTTGTCTAGCTAAAGATTGTAATGACCAATCAGACCTCT 8613  
497 GT-----GTCTAGCTAAAGATTGTAATGACCAATCAGACCTCTGTAATAATGAC 548  
|||  
8612 GTCAAAACGACCAATCAGCTCTGTAAACAGACCAATGCTCTGTGTAATAATGAC 8553  
549 CAATCAGCAGATGTGGGCGGGGTCAATTAAGGAGTAAATATGCGCACCCGAGCAGC 608  
8552 CAATCAGCAGATGTGGGCGGGGTCAATTAAGGAGTAAATATGCGCACCCGAGCAGC 8493  
609 AGTGGCAACCACTCGGGTCCCTTCACACTGTGGAAGCTTTGTTCTTTGCTTTCAC 668  
8492 AGTGGCAACATGCTCCAGTCCCTTCACCAATGTGGAAGCTTTGTTCTTTCACTTAA 8433  
669 AATAATCTTGTGCTGCTCATTTCTTTGTCTCACTACTCTTTAAGAGTGTAAACCTC 728  
|||||  
8432 AATAATCTTGTGCTGCTCATTTCTTTGTCTCACTACTCTTTAAGAGTGTAAACCTC 8373  
729 ACTGGGAGGGTCTGGGCTTCATTCTGAAGTCAAC-AGACCAAGAACCCACTGGAAGGA 787  
8372 ACTGGCAAAAGTCTGTGACTTCTCTCTGAGGCGACAGAACCAACCAAGAGGA 8313  
788 ACAAGAACTCCGACTGTGCTGCTTTAAGAGTGTAAACCTCACTGCGAAGCTCTGAG 847  
8312 ATGAACAAGTCAGATGCGCGCGC-TTCAAGAGATTAACCTCACCGCAAGGCTGCGAG 8254  
848 CTTCACCTCTGAAGTCACTGAGACCAAAACCAACCAAGGAAGAACTCTGACACAC 907  
8253 CTTCACCTCTGAAGTCACTGAGACCAAAACCAACCAAGGAAGAACTCTCAACACAT 8194  
908 CTGAATCTCTGAAGGAACAAATCTCAGACACACATCTTCAGAGCTGTAACTCACTC 967  
8193 CTGAATCTCTGAAGGAACAAATCTCAGACACACATCTTCAGAGCTGTAACTCACTC 8134  
968 CAAGGCTGTGCTTCAATCTTGAAGTCAAGACCAAGAACCAACCAAGGAAGGA 1027  
8133 TGAGGGTCTGAGTTAGTTAGTTGAGTCACTGAGACCAAGAACCAACCAAGGAAGGA 8084  
1028 ATTTCAGACACAGTGAAGAACTGTATTTTGTGCTGCTGAGGTTACTCCAGTC 1087  
8083 ACTGCGACACACAGTGAAGAACTGTATTTTGTGCTGCTGAGGTTACTCCAGTC 8040  
1088 ATTGAAGTCTCAATGAGGCTTAAAGAAACAGAGATGTTTGAAGAGACATGTGG 1147  
8039 -----TTTCATGAGGAGAAAAAAGACG----- 8012  
1148 AATTGTTATGAGACGAGCTTGAAGTCAATGAGGACATTTCTGATCAAACTAGCTGAA 1207  
8011 ----- 8012  
1208 GCAAGGCCAGAAATATTAATTAAGAAACAGTTTGTGAGACAGTATGCTTTTGA 1267  
8011 -----TTTAAACCTCATTTGTAATTACA 7990  
1268 TCTGAGCATGTAGATTATCAAGCAATTAATTAAGAAATATGCGCAGGTGCGATGGCT 1327  
7989 ACAGATATCTTTAAAAATCTCTC-----TGGGCGGACGCGAGTGGCT 7947  
1328 CATGCTGTATCCAGACACTTTTGGAGGCGCAAGGGGTGATCAC--GAGGTCAAGCG 1385  
7946 CAGGCTGTATCCAGACACTTTTGGAGGCGCAAGGGGTGATCAC--GAGGTCAAGCG 7887  
1386 TTGAGACCAAGCTTGGCGCAACAGTGTGAACCCCGTCTTACTAATAAATACAAATTTAG 1445  
7886 TTGAGACCAACCTGGCGCAACAGTGTGAACCCCGTCTTACTAATAAATTTAA 7827

Query 1446 CCTGTGTGTGGCAGCATCTGTAA--TCCAGTACTCAGAGGCTGAGGAGGAAT 1503  
Db 7826 CCAGGTGTGTGTGGCAGCATCTGTAAATCCAGTACTCAGAGGCTGAGTGTAGATGAAT 7767  
Query 1504 CTCTTGAATTTGGAGGCGAGAGTGTGAGTGGCAAGATCAACACAGCACTCATCC 1563  
Db 7766 TCCTTGAACCTTGGAGGCGAGAGTGTGAGTGGCAAGATCAACACAGCACTCATCC 7707  
Query 1564 TGGGTGACAGAGCAGACTCTGTCAAAAAAAGAAAAA 1605  
Db 7706 TGGGCAACAGAGTGTGAGTGTGAGTGGCAAGATCAACAAAAAAGAAAAA 7665

RESULT 28  
US-09-949-016-13572/C  
; Sequence 13572, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13572  
; LENGTH: 80357  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13572

Query Match 19.1%; Score 382; DB 3; Length 80357;  
Best Local Similarity 66.4%; Pred. No. 3.9e-90;  
Matches 785; Conservative 0; Mismatches 210; Indels 187; Gaps 9;

437 GGGGACTTGGAGAACTTTTGTCTAGCTAAAGATTGTAATGACCAATCAGACCTCT 496  
8672 GGGGACTTGGAGAACTTTTGTCTAGCTAAAGATTGTAATGACCAATCAGACCTCT 8613  
497 GT-----GTCTAGCTAAAGATTGTAATGACCAATCAGACCTCTGTAATAATGAC 548  
|||  
8612 GTCAAAACGACCAATCAGCTCTGTAAACAGACCAATGCTCTCTGTAAATAATGAC 8553  
549 CAATCAGCAGATGTGGGCGGGGTCAATTAAGGAGTAAATATGCGCACCCGAGCAGC 608  
8552 CAATCAGCAGATGTGGGCGGGGTCAATTAAGGAGTAAATATGCGCACCCGAGCAGC 8493  
609 AGTGGCAACCACTCGGGTCCCTTCACACTGTGGAAGCTTTGTTCTTTGCTTTCAC 668  
8492 AGTGGCAACATGCTCCAGTCCCTTCACCAATGTGGAAGCTTTGTTCTTTCACTTAA 8433  
669 AATAATCTTGTGCTGCTCATTTCTTTGTCTCACTACTCTTTAAGAGTGTAAACCTC 728  
|||||  
8432 AATAATCTTGTGCTGCTCATTTCTTTGTCTCACTACTCTTTAAGAGTGTAAACCTC 8373  
729 ACTGGGAGGGTCTGGGCTTCATTCTGAAGTCAAC-AGACCAAGAACCCACTGGAAGGA 787  
8372 ACTGGCAAAAGTCTGTGACTTCTCTCTGAGGCGACAGAACCAACCAAGAGGA 8313  
788 ACAAGAACTCCGACTGTGCTGCTTTAAGAGTGTAAACCTCACTGCGAAGCTCTGAG 847  
8312 ATGAACAAGTCAGATGCGCGCGC-TTCAAGAGATTAACCTCACCGCAAGGCTGCGAG 8254  
848 CTTCACCTCTGAAGTCACTGAGACCAAAACCAACCAAGGAAGAACTCTGACACAC 907  
8253 CTTCACCTCTGAAGTCACTGAGACCAAAACCAACCAAGGAAGAACTCTCAACACAT 8194

Oy	308	CTGAATATCTGGAAGGAACAAATCTCAGACACACATCTTTCAAGCTGTAACTCAACG	967
Db	8193	CTGAATCTTGAGAGGAACAAATCTCAGACACATCTTTTAACTGTAACTCAACG	8134
Oy	968	CAAGGCTCTGCTTCATCTTGAAGTCCAGCAAGACCAAGAACCCGAGGAACAA	1027
Db	8133	TGAGGGCTCTGAGGTTTGTCTTGAAGTCAGTGAGACCAAGAACCCACC-----A	8084
Oy	1028	ATTCCACACACAGTAGAAATCTGTATTTTGTATCTGTGCTTCCAGGGTACTCACTC	1087
Db	8083	ACTCCGACACACACAGATTAAAGAAACAAATTTCTGATCTC-----	8040
Oy	1088	ATTGAATCTTCATTTGACGCTTAAAGAAACAGAGATGGTTTGGAGAGACATGTGGG	1147
Db	8039	-----TTTCCATGCGGAAAAAAAAGACG-----	8012
Oy	1148	AATTGTATGAGCACAGGCTTGAGATGCATPAGGGCAATTTCTGATCAACTAGCTGAA	1207
Db	8011	-----	8012
Oy	1208	GCAGGGCCAGAAATATATCTAAGGAAGCAAGTTTTTGTAGACAGTACTTTTGTCA	1267
Db	8011	-----TTTAAACCTCATTTGTAAATTCA	7990
Oy	1268	TCTGAGCATGTAGATTATCAGCAATTAATTAGAAAAATATPAGCAGTGCATGGCT	1327
Db	7989	ACAGATATCTTTAAAAATTTCTTC-----TGGCCGAGCAGATGGCT	7947
Oy	1328	CATGCTGTATATCCACACATTTTGGAGGCAAGGGGTGTGATCAC--GAGTCAAGCTG	1385
Db	7946	CAGCTGTATATCCACACATTTTGGAGGCGCAGAGGGGAGATCACTGAAGTCAAGG	7887
Oy	1386	TTCCAGACCAAGCTCTGGCCAACTGTGTAAACCCCGTCTCTACTAAATATCAAAATTG	1445
Db	7886	TTCCAGACCAACCTGGCCAACTGTGTAAACCCCATCTTTAATAAATCACAAAATTAA	7827
Oy	1446	CCTGGTGTGGGCGACGATCTGTAA--TCCAGTACTCAGGAGGCTGAGGAGGAAAT	1503
Db	7826	CCAGGTGTGGTGTGTTCACCTTAAATCCACGACTCTGGAGAGGCTGTATGATGAAT	7767
Oy	1504	CTCTTGAACCTTGAGGAGCAGAGTTGCATGAGCCAAATCAACACACAGCATCTCATCC	1563
Db	7766	TGCTTGAACCTGGAGGCAAGGTTGAGCGAGCCAAATCATACACTGCATCCAGCC	7707
Oy	1564	TGGGTGACAGAGCAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAA	1605
Db	7706	TGGGCACACAGAGTGAGCTACTCAAAAAAAAAAAAAAAAAAAAAA	7665

RESULT 29

US-09-949-016-184415/c

; Sequence 184415, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREO

; FILE REFERENCE: CUI01307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 184415

; LENGTH: 601

; TYPE: DNA

; ORGANISM: Human

[illegible]

OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12805

```
Query Match      18.6%; Score 372; DB 3; Length 421491;
Best Local Similarity 64.8%; Pred. No. 3.5e-87;
Matches 781; Conservative 0; Mismatches 340; Indels 85; Gaps 12;

QY      494 TCTGTCTAGCTTAAGATTTGTAATGCACCAATCAGCAGCTCTGTAAAT-----543
DB      16909 TTTGTCTAGCTCAGGAGTTGTAACGTACCAATCAGGCGCTGTCAAAACAGAGACT 16850
QY      544 ----TGACCAATCAGCAGATGTGGCGGGGTCAATTAAGGAGTAAATCTGGCCACC 539
DB      16849 GGGCTTACCAATCAGCAGATGTGGGTGGGGCCAGATTAAGAAATTAAGACAGCTGCA 16790
QY      600 CGAGCCAGAGTGGCAACCCACCTCGGGTCCCTTCCACACTGTGGAAGCTTGTCTTTT 659
DB      16789 GGGCCAGAGCGGGCAACCTCTCCGGTCCCTTTCACATTTGGAAGCTTGTCTTTC 16730
QY      660 GCTCTTCACAAATAATCTTGTCTGTCTCATCTTGTGTGCACTACTCTTTATGAGCT 719
DB      16729 GCTGTGTGCAACAACTTGTCTGTCTGTCTCATCTTGTGGTCCACACTGCTTTATGAGCT 16670
QY      720 GTAACTCTCACTCGCAGAGGTCTGTGCTTCACTTCTGAAGTCAAC-AGACCAAGAACCA 778
DB      16669 GTAACTCTCACTCGCAGAGGTCTGTGCTTCACTTCTGAAGTCAACAGACCAAGACCTTA 16610
QY      779 CTGGAAGGAACAAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCACTGCGAA 838
DB      16609 CAAGAGGAACGAACGACTCCACACGCGCCGCTTAAAGAGCTGTAACTCACTCCGGA 16551
QY      839 GCTCTCAGCTTCACTCTGTAAGTCACTGAGACCAACAAACCAAGAGGAAGAACTC 898
DB      16550 GCTCTCAGCTTCACTCTGTAAGTCACTGAGACCAACAAACCAAGAGGAAGAACTC 16492
QY      899 TGAACACCTGTAATCTGAAGGAACAACTCCAGACACACATCTTTCAAGCTGTAA 958
DB      16491 CGAACACATCCGAACATCAAGAGGAACAACTCCAGACACACATCTTTCAAGCTGTAA 16432
QY      959 CATCTCCGGAAGGCTGTGCTTCACTTGTGAAGTCAACCAAGACCAACCAACCCAC-- 1016
DB      16431 CATCTCCGGAAGGCTGTGCTTCACTTGTGAAGTCAACCAAGACCAACCAACCCACAA 16372
QY      1017 -----GGAAGAAACAAATCTCAGACACAGTAAAGAAATCTGTA 1053
DB      16371 TTCCGACACACTCAGAGCCCAAGATGCACTTCAACCAAGAAATGCTCACTTGT 16312
QY      1054 TTTTGAATCTGTGCTTCAAGGCTTACTCAAGTGAAGTCTCACTTGAAGCTTAAAG 1113
DB      16311 ATTGTCAATGCTGTGCTTCAAGGCTTACTTAAAGGGAACCTTGAAGTCTCAACCTTTTCTTCA 16252
QY      1114 GA---AACAAGATGTGTTGAGAGACACTGTGGGAATTTGTAATGAGCAAGCTTGAAG 1170
DB      16251 GGCCTTACATGTGAAGAGATGTTGTCAGTTCATCTTGAAGACAGCAGGCTTTCAGAGC 16192
QY      1171 ATGACATAGGGCA-----TTTCTGATCAAACTAGCTGAAGAGCGGCCAGAAAT 1222
DB      16191 ATGACATAGCTACAGGAGTATGATTTGCTTAAAGCTGGAATTTGAGCTTACAG 16132
QY      1223 ATATCTAAGAGACAGTTTGTGTAGA--CAGTAGTACTTTTGCATCTGAGACATGTA 1280
DB      16131 GATTCATCAAAAGAGCAGCATTTGGAGAGTCTATAGTACAGCACTCAAAAGAGTGTAA 16072
QY      1281 GATTATCAGCAATTAATTAGAAAAATATA-----GCCAGGTGC 1320
DB      16071 GTCAACACAGAGATCCAGGGGTCAACAAGATCAAAAACAGGAGGGGCAACAGCT 16012
QY      1321 GATGCTCATGCTGTATCCAGCACTTTGGAGGGCCAAAG--GGTGTGATCAGCAGG 1378
DB      16011 GGTGCTCATGCTGTATATCCAGCCTTTGGAGAGCAGAGGCGAGAGGTTTCTTGAAGC 15952
QY      1379 TCAGGCTTGGAGACAGGCTTGGCAACATGTGTAAACCCCGTCTTAAATAATACA- 1437
DB      1379 TCAGGCTTGGAGACAGGCTTGGCAACATGTGTAAACCCCGTCTTAAATAATACA- 1437
```

```
DB      15951 CCAGCAGTTTGGAGCACCAGCCGGGTAAATAGAGAGATCTGTCTTCAAAAAATAAT 15892
QY      1438 -----AAATTAAGCTGTGTGTGTGCAAGCATCTGTAATCCAG-TACTCAGAGGCT 1490
DB      15891 AAAAATAAATAATAGCAGGCAATTTGTGACACACTCTTATAGTCCAGACTTTCAGAGCA 15832
QY      1491 GAGGACGGGGAATCTTGTGAATCTGGAGAGCAGAGGTTGCAATGAGCCAAATCACACA 1550
DB      15831 GAGATGGAGGATCTATTGTAGCCCAAGCAGAGTAAAGGCTGAGCTGTGATCATGTCCA 15772
QY      1551 CAGCAGCTCATCTGGGTGACAGAGCAGACTCTGTCTCAAAAAAATAAATAAAG 1610
DB      15771 CTGACTCCAGCATGGGCAACAGAGGAGACTTGTCTCAAAAAAATAAATAATAC 15712
QY      1611 AAAGCA 1616
DB      15711 CAAGCA 15706

RESULT 31
US-09-949-016-14060/c
; Sequence 14060, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14060
; LENGTH: 421494
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(421494)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14060

Query Match      18.6%; Score 372; DB 3; Length 421494;
Best Local Similarity 64.8%; Pred. No. 3.5e-87;
Matches 781; Conservative 0; Mismatches 340; Indels 85; Gaps 12;

QY      494 TCTGTCTAGCTTAAGATTTGTAATGCACCAATCAGCAGCTCTGTAAAT-----543
DB      16909 TTTGTCTAGCTCAGGAGTTGTAACGTACCAATCAGGCGCTGTCAAAACAGAGACT 16850
QY      544 ----TGACCAATCAGCAGATGTGGCGGGGTCAATTAAGGAGTAAATCTGGCCACC 539
DB      16849 GGGCTTACCAATCAGCAGATGTGGGTGGGGCCAGATTAAGAAATTAAGACAGCTGCA 16790
QY      600 CGAGCCAGAGTGGCAACCCACCTCGGGTCCCTTCCACACTGTGGAAGCTTGTCTTTT 659
DB      16789 GGGCCAGAGCGGGCAACCTCTCCGGTCCCTTTCACATTTGGAAGCTTGTCTTTC 16730
QY      660 GCTCTTCACAAATAATCTTGTCTGTCTCATCTTGTGTGCACTACTCTTTATGAGCT 719
DB      16729 GCTGTGTGCAACAACTTGTCTGTCTGTCTCATCTTGTGGTCCACACTGCTTTATGAGCT 16670
QY      720 GTAACTCTCACTCGCAGAGGTCTGTGCTTCACTTCTGAAGTCAAC-AGACCAAGAACCA 778
DB      16669 GTAACTCTCACTCGCAGAGGTCTGTGCTTCACTTCTGAAGTCAACAGACCAAGACCTTA 16610
QY      779 CTGGAAGGAACAAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCACTGCGAA 838
DB      779 CTGGAAGGAACAAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCACTGCGAA 838
```





```

1 FILE REFERENCE: CL001307
2
3 CURRENT APPLICATION NUMBER: US/09/949,016
4
5 CURRENT FILING DATE: 2000-04-14
6
7 PRIOR APPLICATION NUMBER: 60/241,755
8
9 PRIOR FILING DATE: 2000-10-20
10
11 PRIOR APPLICATION NUMBER: 60/237,768
12
13 PRIOR FILING DATE: 2000-10-03
14
15 PRIOR APPLICATION NUMBER: 60/231,498
16
17 PRIOR FILING DATE: 2000-09-08
18
19 NUMBER OF SEQ. ID NOS: 207012
20
21 SOFTWARE: PasteSeq for Windows Version 4.0
22
23 SEQ. ID NO 16105
24
25 LENGTH: 52992
26
27 TYPE: DNA
28
29 ORGANISM: Human
30
31 US-09-949-016-16105

```

Query Match	Similarity	18.5%	Score 370.4	DB 3	Length 52992
Best Local	Similarity	78.3%	Pred. No. 3.7e-87		
Matches	515	Conservative	0	Mismatches	71
				Indels	72
				Gaps	3
Qy	431	TCAGGTGGGAGCTTGGAGAACTTTTGTGTAGCTTAAGAGATTGTAAATGACCAATTCAG	490		
Db	45286	TCAGTGTGGAGACTTGGAGAACTTTTGTGTAGCTTAAGAGATTGTAAATGACCAATTCAG	4523		
Qy	491	CACCTGTGTCTAGCTAAAGGATTGTAAATGC-----	522		
Db	45236	CACCTGTGTCTAGCTCAGGATTGTAAATGCACCACTCAGCACCTGTCAAGACAGACC	45177		
Qy	523	-----ACCAATCAGACCTCTGTAAATATGACCAATTCAGCAGAGT	561		
Db	45176	AAACAGCTCTCTGTAAATATGAGCAACATCCTCTGTAAATATGACCAATTCAGCAGAGT	45111		
Qy	562	GTGGGCGGGGTCAATTAAGAGATTAATAACCTGGCCACCCGAGCAGAGTGGCAACCCAC	621		
Db	45116	GTGGGTGGGGCCAGATTAAAGGAATTAAGACAGGCTGCCCGAGCCAGAGTGGCAACCCAC	45055		
Qy	622	TGGGTGCCCTTCCACATGTGAAAGCTTTGTTCTTTTGTCTCTTCACAATTAATCTTGTCT	681		
Db	45056	TTGGGTGCCCTTCCACATGTGAAAGCTTTGTTCTTTTGTCTCTTGTGCAATTAATCTTGTCT	44999		
Qy	682	GCTGCTCATTTCTTTGTGTGCACTACCTTTATAGCTGTAACTCACTACGTGCGAGGCTCT	741		
Db	44996	GCTGCTCATCTCTTTGGGTGTCACATGCTTTATAGCTGTAACTCACTACCGCGAGGCTCT	44933		
Qy	742	GTGGCTTCAATTCCTGAATGCAACAGACCAAGAACCACTGGAAGAAACAAGAACTGCC	800		
Db	44936	GCAAGTTCACTCTGAGGCGACGACGACCAAGAAATCCACGAGGAGATGAACAATCTCCA	4487		
Qy	801	GATGTGTGCGCT-----TTAAGACTGTAAACATCTGCGAA	838		
Db	44876	GATGGAGAGAAATGACAACTCCAGACGCAACGCTTTAAGACTGTAACTCACCCGTGAA	4481		
Qy	839	GCTCTGACGCTTCACTCTGAAAGTCACTGAGACCAACAACCAACAGAGAGAAATCTC	898		
Db	44816	GCTCTGACGCTTCACTCTGAGAGCAAGGAAACCAACAACCAATCAGAAAGAAATCTC	4475		
Qy	899	TGGACAACACTGAAATATCTGAAGAAACAATCTCAGACAACCATCTTTCAAGACTGTAA	938		
Db	44756	CGAAATATGTCTGAACATCAGAAAGAAACAATCTCGGACACACATCTTTAAGAACTGTAA	4469		
Qy	939	CACCTACCGCAAGGCTGTGTGCTTCAATTTCTTGAAGTACGACAAACCAAGAACCCAC	1016		
Db	44696	CACCTACCGCAAGGCTGTGTGCTTCAATTTCTTGAAGTACGAGAACCAAGAACCCAC	44639		

```

RESULT 34
US-09-949-016-15670/C
; Sequence 15670, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```

```

? TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
? FILE REFERENCE: CLO01307
? CURRENT APPLICATION NUMBER: US/09/949,016
? CURRENT FILING DATE: 2000-04-14
? PRIOR APPLICATION NUMBER: 60/241,755
? PRIOR FILING DATE: 2000-10-20
? PRIOR APPLICATION NUMBER: 60/237,768
? PRIOR FILING DATE: 2000-10-03
? PRIOR APPLICATION NUMBER: 60/231,498
? PRIOR FILING DATE: 2000-09-08
? NUMBER OF SEQ ID NOS: 207012
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 15670
? LENGTH: 33578
? TYPE: DNA
? ORGANISM: Human
US-09-949-016-15670

Query Match      18.4%   Score 367.4; DB 3;   length 33578;
Beet Local Similarity 79.5%   Pred. No. 1.9e-86;
Matches 505; Conservative 0; Mismatches 81; Indels 49; Gaps 4;

```

QY	431	CCAGTGGGGGACCTTGAGAACTTTTGGTCTAGCTGAAGAATTGTAATGACCAATACAG	490
Db	7467	TCTAGTGGGGACCTTGAGAACTTTTATGTCTAGCTGAAGATTGTAATGACCAATACAG	7408
QY	491	CACCTGTGTCTAGCTAAAGGATTGTAAATGCAACCAATCAGCA-----	533
Db	7407	CATTCTGTCTAGCTCAGGGATTGTAAACGCAACCAATCAGCACTCTGTCAAAACGAC	7348
QY	534	-----CTCTGTAAATTGACCAATCAGAGATGTGGGCGGGGTCAATTAAGGGAGT	585
Db	7347	AATCAGCTCTCTGTAAATTGACCAATCAGAGATGTGGGCGGGGTCAATTAAGGGAGT	7288
QY	586	AAAACTGGACCCGAGCCGAGAGTGGGCAACCACTGGAGTCCCTTCCACACTGTGGA	645
Db	7287	AAAGCAGAGCTGCCCAAGCCAGACGCGGCACACTGCTGGGTCCCTTTGACACTGTGGA	7228
QY	646	AGCTTGTCTTTTGTCTCTTCACAATAAATCTTGCTGTGCT -CATTCCTTTGTCTCACA	704
Db	7227	AGCTTGTCTTTTCAAGTCTTTGCAATAAATCTTGCTGTCTCTCACTCGGATCGGCA	7168
QY	705	CTACCTTATAGCGCTGAACACTACGTCGAGGGGTGTGGCTTCATCTCGAAGTCAAC	764
Db	7167	CTGCCGTATAGCTGTAACTACACTCTCGAAGGTCTGCAGCTTCACTCTGAGGCCAGC	7108
QY	765	-AGACCAAGAACCCCACTGGAAGGAACAAGAACTCCGATGTCTGC-----	810
Db	7107	GAGACCAAGAACCCCACTGGAAGGAATGAACAATCTCAACAGAGGAACCTAAACAATCCA	7048
QY	811	-----CTTTAAGACTGTAAACACTCACTGCGAAGCTCTGCAAGCTTCACTCCTGAA	861
Db	7047	GATGCGCGCCCTTTAAGACTGTAAACTGTACCACTGACCAAGAGTCTGCAGCTTCACTCCTGAA	6988
QY	862	TCAGTGAAGACCAACAACCCACAGAGAAGAAAGAACTCTGGAACAACACTGAATATCTGAAG	921
Db	6987	CCAGCAAGACCAACAACCCACAGAGAAGAAAGAACTCCGAACATGTCCGAACATCAGAAG	6928
QY	922	GAAACAACTCGAGACACACATCTTTCAAGCTGTAACTCACTCAACGCAAGGTCTGTGGC	981
Db	6927	AAACAAACTCCGAGACACACATCTTTAAGAACTGTAAACTCACTCACTCAAGGTCGCGGC	6866
QY	982	TTTCATCTTGAAGTCAGAGACCAAGAACCACC 1016	
Db	6867	TTTCATCTTGAAGTCAGAGACCAAGAACCACC 6833	

RESULT 35  
US-09-949-016-16041/c  
; Sequence 16041, Application US/09949016  
; Patent No. 681239  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.



```

? TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
? TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
? FILE REFERENCE: C1001307
? CURRENT APPLICATION NUMBER: US/09/949,016
? CURRENT FILING DATE: 2000-04-14
? PRIOR APPLICATION NUMBER: 60/241,755
? PRIOR FILING DATE: 2000-10-20
? PRIOR APPLICATION NUMBER: 60/237,768
? PRIOR FILING DATE: 2000-10-03
? PRIOR APPLICATION NUMBER: 60/231,498
? PRIOR FILING DATE: 2000-09-08
? NUMBER OF SEO ID NOS: 207012
? SOFTWARE: FastSeq for Windows Version 4.0
? SEO ID NO: 16041
? LENGTH: 87644
? TYPE: DNA
? ORGANISM: Human
? US-09-949-016-16041

```

```

Query March 18.3%; Score 365.4; DB 3; Length 87644;
Best Local Similarity 82.6%; Pred. No. 9.8e-86;
Matches 493; Conservative 0; Mismatches 91; Indels 13; Gaps 6;

QY 431 TCAGTGGGGGCACTTGAGAACTTTGTGTCTAGCTAAAGATTGTAATGACCAATCAG 490
Db 34681 TCTAGTGGTGAACGTGGAGAACTTTGTGTCTAGCTACGGAGTTGTAAACGCAATCAG 34622
QY 491 CACTCTGT-----GCTTAGCTAAAGAGATTGTAATGACCAATCAGCACTCTGTAA 542
Db 34621 CACCGTGTCAAAATATGACCAATCAGCCCTCTGTAAACAGACCAATCGGCTCTGTAA 34562
QY 543 ATGACCAATTCAGAGATGTGGGGGGGGTCAATTAAGGAGTAAATACGGCCACCCGA 602
Db 34561 ATGGGCCAATTCAGAGATGTGGGTGGGGCCAGATTAAGAAATATTAAGCAGGCTGCCGA 34502
QY 603 GCCAGCATGGGCAACCACTCGGGTCCCTTCCACACTGTGGAAGCTTTGTCTTTTGTCT 662
Db 34501 GCCAGCATGGGCAACCCGCTCGGTCCTCCCTTCCACTACTGTGGAAGCTTTGTCTTTTGTCT 34442
QY 663 CTTGACAAATAATCTTGTCTGTCTCACTTTTGTGTCAACTACCTTTATAGCTGTA 722
Db 34441 CTTGCAATAAATCTTGTCTGTCTCACTTTTGTGTCAACTACCTTTATAGCTGTA 34382
QY 723 ACACTCACTCGGAGGGTCTGTGGCTTCACTTCTGAATCAACAGACCAAGACCCACTG 781
Db 34381 ACACTCACTCGGAGGGTCTGTGGCTTCACTTCTGAATCAACAGACCAAGACCCACTG 34322
QY 782 GAGGGAACAAGAACTCCGATGTGCTGCTTTAAGAGCTGTAACTCACTCGGGAAGCT 841
Db 34321 GAGGGAACAAACGACTCCAGCTCACTGCTCC-TTAAAGAGCTGTAACTCACTCGGGAAGCT 34263
QY 842 CTGCAGCTTCACTCTTGAAGTCACTGAGACCAACCAACCAAGAAAGAACTGTG 901
Db 34262 CTGCAGCTTCACTCTG-AGCAGTAAAGACATTAACCAACCAAGAAAGAACTTGTG 34204
QY 902 ACACACCTGAATATCTGAAGAACAAACTCCAGACACACATCTTTCAGAGCTG-TTACA 960
Db 34203 ACACATCCCAATCAGAGAAACAAACTCTGACATGCGGCTTTAAGAACTGTTAACA 34144
QY 961 CTCACCCGCAAGGGTCTGTGGCTTCACTTCTT-GAAGTCAGCAAGCAAGAACCCAGC 1016
Db 34143 CTCACCCGCAAGGGTCCGTGGCTTCACTTCTTGAAGTCAATGAGCAAGAACCCAGC 34087

RESULT 36
US-09-422-576D-6
; Sequence 6, Application US/09422576D
; Patent No. 6395549
; GENERAL INFORMATION:
; APPLICANT: Tuan, Dorothy
; APPLICANT: Benga, Oiaoming
; APPLICANT: Benga, Chik
; TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use in

```

```

: TITLE OF INVENTION: Recombinant Vectors
:
: FILE REFERENCE: M0351-205010
:
: CURRENT APPLICATION NUMBER: US/09/422,576D
:
: CURRENT FILING DATE: 1999-10-21
:
: PRIOR APPLICATION NUMBER: US 60/105,256
:
: PRIOR FILING DATE: 1998-10-22
:
: NUMBER OF SEQ ID NOS: 25
:
: SOFTWARE: Patentin version 3.1
:
: SEQ ID NO 6
:
: LENGTH: 1043
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: US-09-422-576D-6

```

Query Match	18.1%	Score 361.8	DB 3	Length 1043
Best Local Similarity	80.5%	Pred. No. 1,2e-85		
Matches 466	Conservative 0	Mismatches 97	Indels 16	Gaps 3
QY	453	TTTGTCTCTAGCTAAAGATTGTAAATGCAACCAATCAGCACTCTGTGTCTTAGCTAAAGGA	512	
Db	359	TCGTATCTAGCTCAAGGTTTGTAAACACCAATCAGCACTCTGTGTCTTAGCTCAAGGA	418	
QY	513	TTGTAAATGCACCATTCAGCACTCTGTAAATGCAACCAATCAGCACTCTGTGTCTTAGCTCAAGGA	558	
Db	419	TTGTAAATGCACCATTCAGCACTCTGTAAATGCAACCAATCAGCACTCTGTGTCTTAGCTCAAGGA	478	
QY	559	GATGTGGGCGGGGCTCAAAATTAAGGAGTAAAACTGGCCACCAGCCAGCAAGTGGCAAC	618	
Db	479	GATGTGGGCGGGGCTCAAAATTAAGGAGTAAAACTGGCCACCAGCCAGCAAGTGGCAAC	538	
QY	619	CACGTGGGCGGGGCTCAAAATTAAGGAGTAAAACTGGCCACCAGCCAGCAAGTGGCAAC	678	
Db	539	CGGTGGGCGGGGCTCAAAATTAAGGAGTAAAACTGGCCACCAGCCAGCAAGTGGCAAC	598	
QY	679	GCTGTGCTCATCTCTTGTGTGTCACACTACCTTTATGAGCTGTAACTCACTGCGAGG	738	
Db	599	GCTGTGCTCATCTCTTGTGTGTCACACTACCTTTATGAGCTGTAACTCACTGCGAGG	658	
QY	739	TCTGTGGCTTCACTTCTGTGAAGTCAACAGACCAAGCCACTGGAAGAACT	797	
Db	659	TCTGTGGCTTCACTTGTGAAGTCAACAGACCAAGCCACTGGAAGAACT	718	
QY	798	CCGATGTGCTGCTTAAAGCTGTAACTCACTGCGAGGCTCTGCACTTCACTCT	857	
Db	719	CCGATGTGCTGCTTAAAGCTGTAACTCACTGCGAGGCTCTGCACTTCACTCT	778	
QY	858	GAGTCAGTGAAGCACAACCCACAGAGGAAGAACTCTGGACAACCTGATATCT	917	
Db	779	GAGTCAGTGAAGCACAACCCACAGAGGAAGAACTCTGGACAACCTGATATCT	837	
QY	918	GAGGAAACAATCCAGACACCACTTTTGAAGCTGTAACTCACTGCGAGGCTG	977	
Db	838	GAGGAAACAATCCAGACACCACTTTTGAAGCTGTAACTCACTGCGAGGCTG	897	
QY	978	TGGCTTCATCTTGAAGTCAAGACCAAGACCAACCCACC	1016	
Db	898	TGGCTTCATCTTGAAGTCAAGACCAAGACCAACCCACC	936	

```
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11817
/ LENGTH: 124700
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-11817

Query Match
Best Local Similarity 18.0%; Score 359.8; DB 3; Length 124700;
Pred. No. 3,4e-84;
Matches 469; Conservative 0; Mismatches 87; Indels 27; Gaps 3;

QY 464 CTAAAGATTGTAATGACCAATCAGCACTGTGTCTAGCTAAAGATTGTAATGCA 523
DB 30935 CTGTAAATGAGACCAATCAGCTCTGTAAAGAACCAATCAGCACTGTAAATGGA 30876
QY 524 CCATCAGCACTCTGTAAATGAGACCAATCAGCAAGATGTGGCGGGGTCAATTAAGGA 583
DB 30875 CCATCAGCTCTGTGAAATGAGACCAATCAGCAAGATGTGGCGGGGTCAATTAAGGA 30816
QY 584 GTAAATGAGCCACCGGACGAGCAGTGGCAACCACTCGGGTCCCTTCCACTGTG 643
DB 30815 ATAAAGCTGGCCACCTGAGCCAGCGGCAACCCGCTCAGATCCCTTCCACTGTG 30756
QY 644 GAAGCTTTGTTCTTTTGTCTTCAATTAATCTGTGCTCATCTTTGTGTCCAC 703
DB 30755 GAATCTTTGTTCTTTGCTCTTCAATTAATCTGTGCTCATCTTTGGGCTTGC 30696
QY 704 ACTACCTTATGAGCTGTAACTCACTCAGGAGGTCTGTGGCTTCACTCTGAAGTCA- 762
DB 30695 ACTACCTTATGAGCTGTAACTCACTCAGGAGGTCTGTGGCTTCACTCTGAAGTCA 30636
QY 763 ACAGACCAAGAACCCCACTGAGAGAACAAAGAACTCCGATGTGCTCTTAAAGCTG 822
DB 30635 TGAGACCAAGAACCCCACTGAGAGAACAAAGAACTCTGATGGCCACTTTAAAGCTG 30576
QY 823 TAACACTCACTGAGAGCTCTGAGCTTCACTCTCTGAAGTCAAGTGAACCAACCCAC 882
DB 30575 TAACACTCACTGAGAGCTTGGCGCTTCACTCTCTGAAGTCAAGTGAACCAACCCAC 30516
QY 883 CAGAAGAGAAACTCTGACACACCTGAATATCTGAAGAAACAATCCAGACACCA 942
DB 30515 TGAAGAGAAACTCTGAGACAC-----ATCTGAAGAAACAATCTGAGACACCA 30464
QY 943 TCTTCAGAGCTGTAACTCACTCAGCAAGGCTGTGGCTTCACTTTGAAGTCAAG 1002
DB 30463 TCTTTAAGAACTGTAACTCACTCAGCAAGGCTC-----CACAGTGAAG 30422
QY 1003 ACCAAGAACCCACCGAAGAGAACAAATTCAGACACAGTGA 1045
DB 30421 ACCAAGAACCCACCGAATGAATTAATTCGAGACACAAAGAA 30379

RESULT 38
US-09-949-016-15439/c
/ Sequence 15439, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
```

```
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15439
/ LENGTH: 124701
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-15439

Query Match
Best Local Similarity 18.0%; Score 359.8; DB 3; Length 124701;
Pred. No. 3,4e-84;
Matches 469; Conservative 0; Mismatches 87; Indels 27; Gaps 3;

QY 464 CTAAAGATTGTAATGACCAATCAGCACTGTGTCTAGCTAAAGATTGTAATGCA 523
DB 30935 CTGTAAATGAGACCAATCAGCTCTGTAAAGAACCAATCAGCACTGTAAATGGA 30876
QY 524 CCATCAGCACTCTGTAAATGAGACCAATCAGCAAGATGTGGCGGGGTCAATTAAGGA 583
DB 30875 CCATCAGCTCTGTGAAATGAGACCAATCAGCAAGATGTGGCGGGGTCAATTAAGGA 30816
QY 584 GTAAATGAGCCACCGGACGAGCAGTGGCAACCACTCGGGTCCCTTCCACTGTG 643
DB 30815 ATAAAGCTGGCCACCTGAGCCAGCGGCAACCCGCTCAGATCCCTTCCACTGTG 30756
QY 644 GAAGCTTTGTTCTTTTGTCTTCAATTAATCTGTGCTCATCTTTGTGTCCAC 703
DB 30755 GAATCTTTGTTCTTTGCTCTTCAATTAATCTGTGCTCATCTTTGGGCTTGC 30696
QY 704 ACTACCTTATGAGCTGTAACTCACTCAGGAGGTCTGTGGCTTCACTCTGAAGTCA- 762
DB 30695 ACTACCTTATGAGCTGTAACTCACTCAGGAGGTCTGTGGCTTCACTCTGAAGTCA 30636
QY 763 ACAGACCAAGAACCCCACTGAGAGAACAAAGAACTCCGATGTGCTCTTAAAGCTG 822
DB 30635 TGAGACCAAGAACCCCACTGAGAGAACAAAGAACTCTGATGGCCACTTTAAAGCTG 30576
QY 823 TAACACTCACTGAGAGCTCTGAGCTTCACTCTCTGAAGTCAAGTGAACCAACCCAC 882
DB 30575 TAACACTCACTGAGAGCTTGGCGCTTCACTCTCTGAAGTCAAGTGAACCAACCCAC 30516
QY 883 CAGAAGAGAAACTCTGACACACCTGAATATCTGAAGAAACAATCCAGACACCA 942
DB 30515 TGAAGAGAAACTCTGAGACAC-----ATCTGAAGAAACAATCTGAGACACCA 30464
QY 943 TCTTCAGAGCTGTAACTCACTCAGCAAGGCTGTGGCTTCACTTTGAAGTCAAG 1002
DB 30463 TCTTTAAGAACTGTAACTCACTCAGCAAGGCTC-----CACAGTGAAG 30422
QY 1003 ACCAAGAACCCACCGAAGAGAACAAATTCAGACACAGTGA 1045
DB 30421 ACCAAGAACCCACCGAATGAATTAATTCGAGACACAAAGAA 30379

RESULT 39
US-09-949-016-12372/c
/ Sequence 12372, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 12372
```

```

; LENGTH: 181429
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12372

Query Match      17.9%; Score 357.2; DB 3; Length 181429;
Best Local Similarity 81.8%; Pred. No.26-83;
Matches 481; Conservative 0; Mismatches 83; Indels 24; Gaps 5;

OY      431 TCAGGTGGGAGCTTGGAGAACTTTGTGTCTAGCTAAAGATTGTAATGCAACCAATCAG 490
Db      138282 TCTAGTGGGAGCATGGAGAACTTTGTGTCTAGCTCAGGGATTGTAAACACCAATCAG 138223
OY      491 CACTCTGTCTTAGCTAAAGATTGTAATGCAACCAATCAGCACTCTGTAAATGAGCA 550
Db      138222 CACCCGTCA-----AAACAGCCATCAGGCTCTCTGTAAACACAGCCA 138179
OY      551 ATCAGAGGATGTGGGCGGGGTCAATAATAGGAGTAAATACTGGCCACCCGAGCCAGCAG 610
Db      138178 ATCAGAGGATGTGGGTGGGCGCATTAAGAGATTAAGACAGGCTGCCGAGCCAGCAG 138119
OY      611 TGGCAACCCACTCGGGTCCCTTCCTCCACACTGTGGAAGCTTTGTCTTTTGTCTTCACAA 670
Db      138118 TGGCAATCCTCTCAGGTCCCTTCCTCCACACTGTGGAAGCTTT-----TTCACTTTTGTGA 138064
OY      671 TAAATCTGTCTGTCTGTCAATTC-TTTGTGTCCACACTACCTTTTGTAGCTGTAACTCA 729
Db      138063 TAAATCTGTCTGTCTGTCTCACTCTTTTGGGTCCACATTTGCTTTTAGCTGTAACTCA 138004
OY      730 CTGGCAGAGGTCTGTGGCTTCATTTCTGGAAGTCAAC-AGACACAGAACCCCACTGGAAGGAA 788
Db      138003 ATGCCAAGGCTGTGCAGCTTCACCTCTGGAAGCCAGCAAGACCAACCAACGAAAGGAA 137944
OY      789 CAAAGAACTCCCGAGTGTGTGCTCTTAAGAGCTGTAACTACATGTCGAAAGCTCTGCAAG 848
Db      137943 CGAAACAATCCAGAGCACTGCTCC-TTAAAGCTGTAACTACACCGGAAAGGCTTCTCAGC 137885
OY      849 TTCACCTCTGAAGTCAGTGAGACCAACAAACCCACGAAAGAAAGAACTCTGACACAAC 908
Db      137884 TTCACCTCCGAGCCAGCGAGACACGAGACCCACGAAAGAAAGAACTCCAAACACATC 137825
OY      909 TGAATATCTGAAGAAACAAACTCCAGACACACCATTTTTCAGAGCTGTAACTCACTCCGC 968
Db      137824 CGAAATATCAGAAAGAAACAAACTCTGACACGCACTTTTAAAACTGTAACTCACTCAC 137765
OY      969 AAGGCTGTGTGCTCATTTCTTGAAGTCAGCAAGACCAAGAACCCAC 1016
Db      137764 GAGGCTCCGCGGCTTCATTTCTTGAAGTCAGTGAGACCAAGAACCCCC 137717

RESULT 40
US-09-949-016-15772/c
; Sequence 15772, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15772
; LENGTH: 181430
; TYPE: DNA
; ORGANISM: Human

```

```
US-09-949-016-15772
Query Match      17.9%; Score 357.2; DB 3; Length 181430;
Best Local Similarity 81.8%; Pred. No. 2e-83;
Matches 481; Conservative 0; Mismatches 83; Indels 24; Gaps 5;

QY      431   TCAAGTGGGGAATTGAGAACCTTTGTGTCTAAGCTAAAGATTTGTAATGACCCATTGCG 490
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       138282 TCATATGGGAGCATGAGAAACTTTTGTTGTCTAAGCTCAGGATTTGTAACAACCAATCAG 138223
QY      491   CACTGTGTCTAAGCTAAAGAATTGTAATGACCACATGAGCACTGTGTAATAATGAGCCA 550
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       138222 CACCCTGTCA-----AAAAGACCAATAGCTCTCTGTAAAAACAGACC 138179
QY      551   ATCAGCAGATGTGGGCGGGGTCAAATPAAAGGAGTAAAAACTGGGCCACCCGAGCCAGAG 610
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       138178 ATCAGCAGATGTGGGTGGGGCCAGATPAAAGATPAAAGCAGGCTGCCGAGCCAGCAG 138119
QY      611   TGGCAACCACTCGGGGTCCCCTTCACACACTGTGGAAGCTTTGTTCTTTGCTTACAA 670
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       138118 TGGCAATCCTCTCAGAGTCCCCCTTCACACTGTGGAAGCTTTTCACTCTTTGTGA 138064
QY      671   TAAATCTTGCTGCTGCCTCAATTC-TTGTGTCCACCTACCTTATGAGGTGTAACTCA 729
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       138063 TAAATCTTGCTGCTGCCTCACTTTTGTGGTTCACATGTGCTTTATGAGGTGTAACTCA 138004
QY      730   CTGGCAGAGGTCTGTGGCTTCAATTCCTGAAGTCAAC-AGACACAGAACCCACTGSAAGAA 788
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       138003 ATGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGACAGAACCCACAGAAAGAA 137944
QY      789   CAAGAACCTCCGATGTGTCTGCTTTAAAGCTGTAACTCACTGCGAAGCTCTTGACG 848
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       137943 CGAACCACTCCAGAGCACTGCGC-TTAAAGCTGTAACTCACCGCGAAGGTCTTGACG 137885
QY      849   TTCACTCCTGGAAGTGAAGTGAACCAACAACCCACCAAGAAAGAAACTCTGACACACG 908
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       137884 TTCCTCTCCGAAAGCCAGGAGCAACGAGGCCACCAAGAAAGAAACTCCAAACATC 137825
QY      909   TGAATATCTGAAGGAACAACCTCCAGACACACCATCTTTCAAGAGCTGTAACTCACCGG 968
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       137824 CGAATATGAGAAAGAAACAACCTCTGGAACGCGCACTTTAAATACTGTAACTCACCG 137765
QY      969   AAGGCTCTGTGCTTCAATTTTGAAGTCAGCAAGCAAGAACCCAGC 1016
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       137764 GAGGTCGCGGCTTCAATTTGAAGTCAGTGAAGCAAGAACCCCGC 137717

RESULT 41
US-09-949-016-13203
; Sequence 13203, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq For Windows Version 4.0
SEQ ID NO 13203
LENGTH: 94019
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(94019)
```

OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13203

Query Match 17.7%; Score 355; DB 3; Length 94019;  
Best Local Similarity 60.4%; Pred. No. 5.6e-83;  
Matches 889; Conservative 0; Mismatches 425; Indels 159; Gaps 12;

QY 409 TGAAGCAGCTGGGCTTCTGGGTCAAGTGGGGAATTGGAGAACTTTGTCTAGCTTAA 468  
DB 88223 TGTATCTACTGCTCTGTGGGGCTTGGAGAACGTGTGTGGAACTCTGTATTAAC 88282  
QY 469 GGATTGTAAATGACCAATCAGACTCTGTGTCTAGCTTAAAGGATTGTAAATGACCAAT 528  
DB 88283 TATCTGATGGGGAAGCTGGAGAACCTTTGTATCTACTCTCAGGATTGTAAATGTACCAAT 88342  
QY 529 CAGCACTCTGTATAA-----TGCACATACGAGAGATGTGGGGGGGTCA 574  
DB 88343 CAGCGCCCTGACAAAACAGGCCACTGGGCTTACCAATCAGAGATGTGGGGGGGTCA 88402  
QY 575 AATTAAGAGATAAATACTGGCCACCGAGCAGAGTGGCAACCACTCGGGTCCCTTC 634  
DB 88403 GATTAAGATATAAAGCAGGCTCCGAGCCCGCATTTGGCAACCACTCGGGTCCCTTC 88462  
QY 635 CACACTGTGAGAGCTTTGTTCTTTTGTCTTCACAATAATCTTGTCTGTCAATTT 634  
DB 88463 CACACTGTGAGAGCTTTGTTCTTTTGTCTTCAATAATCTTGTCTGTCAATTT 88522  
QY 695 TGTGTCCACACTACCTTATAGCTGTAACTCACTCAGGAGGTCTGTGCTTCAATTC 754  
DB 88523 TGGGTCCACGCTCTTTATGAGTTGTAACTCACTCAGGAGGTCTGTGCTTCAATTC 88582  
QY 755 TGAAGTCAAC-AGACCAAGAACCCACTGGAAGAAACAAGAACTCCGATGTGTGCTT 813  
DB 88583 TGAACCCAGGAGAACCAAGAACCCACTGGAAGAAACAAGAACTCCGATGTGTGCTT 88641  
QY 814 TAAAGCTGTAACTCACTGTGCAAGCTCTGCACTTCACTCTGAAATGATGAGACCA 873  
DB 88642 TAAAGCTGTAACTCACTGTGCAAGCTCTGCACTTCACTCTGAA-AGCCAGCAGACCA 88700  
QY 874 CAATCCACAGAAAGAAAGAACTCTGCAACCTGAAATATCTGAAAGAAACAATCCCA 933  
DB 88701 TGAACCCACAGAAAGAAAGAACTCTGCAACCTGAAATATCTGAAAGAAACAATCCCA 88760  
QY 934 GACACACATCTTTCAGAGCTGTAACTCACTCAGCAAGGCTGTGTGCTTCACTTCTTGA 993  
DB 88761 GATGTGCCA-CTTTAAGCTGTAA-----CAGCGCAGGGTCCCGGCTTCACTTGA 88815  
QY 994 GTGACGAAGCAAGAACCAAC-----G 1017  
DB 88816 GTGACGAAGCAAGAACCAACCAATTCGGGACACCAATACGTGATAGGATAGATC 88875  
QY 1018 GAAGGAACAATTCACACACAGTGAAGAAATCTGTATTTTGTGCTGTGCTTCCAGG- 1076  
DB 88876 AAGGGGAATCAAACTAAGATGAGAAAGCCAGTAAAGGCACTGCAATTCACAGG 88935  
QY 1077 -----TTACTCCAGTCAATTAAGTCTCCATTTGAGCCTTAAGAAACAG 1120  
DB 88936 AAGAGATAGAGCTTAATTAATTAATGATCCAGTGCATTAAGAGAGAGGAGCAGC 88995  
QY 1121 AGAATGTTTGGAGAG-----CACATGTGGAAATTTGTTATGACCAAGCTTGAATG 1173  
DB 88996 CAGGTGAGAGTGAAGAGAGAGCAAGCAGGCTGGGTGCTGAAATTCAGTTGCAAGC 89055  
QY 1174 CACATAGGGCACTTCTGTATCAACCTAGCTGAAAGCAGGAGCAGAGAAATATATCTAAG 1233  
DB 89056 TGGCAGATGTAGTGTGTGTATCACTCAACGAAAGGAGAGCCAGAAAGAGTG 89115  
QY 1234 AAGCAGATTTTGTAGACAGTATGCTTGTGATCTGAGCATGTATATCAAGCA 1293  
DB 89116 ATTGTGAGAAACGTGGGGAGCTTATGTTCTGTGTCTTTGATTTTGAATACTGGGT 89175  
QY 1294 TTAATTGAAAAATA----- 1309

DB 89176 TTGAAGAGAACCAATTAATTAATGATCTAGTGAATATGCTTGTGCTTACAGAAAG 89235  
QY 1310 -----TACCAGGTGCGATGCTCAATGCTTATTC 1340  
DB 89236 ATCAGAACCTTTGATTAAGAACCTTGCCCTGCTGGCGCGGTGCTCAGCCTGTATTC 89295  
QY 1341 CCAGCATTTGGGAGCCCAAGGGGTGTGATACAGAGTCAAGCGTTTCAAGCAGCCCTG 1400  
DB 89296 CCAGCATTTGAGAGCCCAAGGGGTGTGATACAGAGTCAAGAGTCAAGAGATCAAGCAATCCTG 89355  
QY 1401 GCCAATGATGAAACCCCGCTCTACTTAAATAA-----CAAAATTAAGCTGTGTGTG 1457  
DB 89356 GCTAACAGGTAAACCCAGTCTTACTTAAATAATCAAAATAATTAAGCTGTGTGTG 89415  
QY 1458 GCAGCATCTGTATCTCCAG-TACTCAGAGCTGAGGAGGGAATCTTGAATTG 1516  
DB 89416 GCGGAGCTGTATGTCCAGCTACTGCGGAACTGAGGAGATGCGGGAACCCG 89475  
QY 1517 GAGCAGAGTTGCACTGAGCCAGATCACACAGCACTCTCATCTGTGGTGAAGAGC 1576  
DB 89476 GAGCGAGCTTGTGTAAAGCCAGATCACGCACTGCACTCAGCTGTGGTGAAGAGC 89535  
QY 1577 GAGCTGTGTCAAAAAAAGAAAGAAAGAAATATATCAAGATATG 1636  
DB 89536 GAGCTGTGTCAAAAAAAGAAAGAAAGAAATATATCAAGATATG 89595  
QY 1637 ACAGTAACTTATTAATCAACTTACTATGACACAGCAATACATAAGTGTATTAATG 1696  
DB 89596 AATAGAGCTGTATTAATCAACTTACTATGACACAGCAATACATAAGTGTATTAATG 89655  
QY 1697 GATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1729  
DB 89656 TGTGATCTGTGCTTAAATTAATGATGATAT 89688

RESULT 42  
US-09-949-016-17028/c  
; Sequence 17028, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTNER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ. ID NOS: 207012  
; SOFTWARE: PatsSeq for Windows Version 4.0  
; SEQ. ID NO 17028  
; LENGTH: 88557  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(88557)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-17028

Query Match 17.5%; Score 350.2; DB 3; Length 88557;  
Best Local Similarity 79.3%; Pred. No. 1e-81;  
Matches 488; Conservative 0; Mismatches 98; Indels 29; Gaps 5;

QY 431 TCAAGTGGAGACTTGGAGAACTTTGTCTAGCTTAAAGATTTGAATGACCAATCAAG 490  
DB 78270 TCTAGTGGGACCTTGGAGAACTTTGTCTAGCTCAGATTTGAAGCAACCAATCAAG 78211  
QY 491 CACTGTGT-----GTCTAGCTTAAAGATTTGAATGACCAATCAAGCACTCTGTAATAATG 546

Db 78210 CACCTGTCAAGACGACCAATGAGCTCTGTGTAACCAATGAGCTCTGTGTAATATGC 78151  
QY 547 ACCAATCAGAGAGATGTGGGCGGGGTCAATTAAGGAGTAAAA-CTGGCACCAGGCC 605  
Db 78150 ACCAATCAGTAATGTGGGTGGGGCCAGATTAAGGATTAAGGAGGCTGCGAGGCC 78091  
QY 606 AGCAGTGGCAACCACTCGGGGTCCCTTTCACACTGTGGAAGCTTGTCTTTTGTCTT 665  
Db 78090 AGCAGTGGCAACCTGCTCAGGCCCCCTTTCACACTGTGGAAGCTTGTCTTTTGTCTT 78031  
QY 666 CACAATTAATCTTGTGCTGCTCACTTCTTGTGTCACTCACTCACTTATGAGCTGTACA 725  
Db 78030 TGCATTAATCTTGTGCTGCTCACTTCTTGTGTGTGCGCACTGCTTTGTGAGCTGTACA 77971  
QY 726 CTCACTGCGAGGCTGTGCTTCTTCTTCTTCTGAACTCAAC-AGACCAAGAACCACTGGA 784  
Db 77970 CTCACGCGGAAGGTGTGCACTCTTCACTGAGGCGCAAAACCAAGAACCACTGGA 77911  
QY 785 GGAACAAAGAACTCCGATGTGCTG-CTTTAAGAGCTG 822  
Db 77910 GGAATTAACACTCTCGAGCGGAGAAAGAAAGCACTGCTCAGCTGCTGCTTTAAGAGCTG 77851  
QY 823 TAACACTCACTGCGAAGCTTGTGAGCTTCACTCTGAAAGTCACTGAGAACCAAACTGAG 882  
Db 77850 TAACACTCACTGCGAAGGTGTGAGCTTCACTCTGAAAGTCACTGAGAACCAAACTGAG 77791  
QY 883 CAGAGAGAGAACTCTGAGACACCTGAAATATCTGAGAGAACAACTCCAGACACCA 942  
Db 77790 CAGAGAGAGAACTCTGAGACACCTGAAATATCTGAGAGAACAACTCCAGATACCA 77731  
QY 943 TCTTTAGAGCTG-TAACACTACCGCAAGGCTGTGGTCTTATTTTAATCTGAGCA 1001  
Db 77730 TCTTTAGAGCTGTAACTCACTGAGGAGGTCGCGGCTTCTTGAAGTCACTGTA 77671  
QY 1002 GACCAAGAACCCACC 1016  
Db 77670 GACCAAGAACCCACC 77656

RESULT 43  
US-09-949-016-16179  
; Sequence 16179, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16179  
; LENGTH: 157644  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(157644)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16179

Query Match 17.5%; Score 349.2; DB 3; Length 157644;  
Best Local Similarity 82.7%; Pred. No. 2,4e-81;  
Matches 435; Conservative 0; Mismatches 88; Indels 3; Gaps 3;

QY 517 AATGACCAATCAGACTCTGTAAATGAGCAATCAGAGATGTGGCGGGTCAAA 576  
Db 31518 AATGACCAATCAGACTCTGTAAATGAGCAATCAGAGATGTGGCGGGTCAAA 31577  
QY 577 TAAGGAGTAAAACTGGCCACCCGACGACGAGTGGCAACCACTCGGGTCCCTTCA 636  
Db 31578 TAAGGAGTAAAACTGGCCACCCGACGACGAGTGGCAACCACTCGGGTCCCTTCA 31637  
QY 637 CACTGTGGAAGCTTGTCTTGTCTTGTCTTCAATTAATCTGCTGCTCACTTCTG 696  
Db 31638 TGTGTGGAAGCTTGTCTTGTCTTGTCTTCAATTAATCTGCTGCTCACTTCTG 31697  
QY 697 TGTCACTACTCTTATGAGCTGTAACTCACTGAGAGGTGTGAGCTTCACTTCTG 756  
Db 31698 GGTTCGCTGCTTATGAGCTGTAACTCACTGAGAGGTGTGAGCTTCACTTCTG 31757  
QY 757 AAGTCAAC-AGACCAAGAACCACTGGAAGAAAGAAAGAAAGAAAGTGTGCTGCTT 815  
Db 31758 AAGTCAACAGAACTGAACTGCTGCGGAGAAAGAAAGAAAGTGTGCTGCTG 31817  
QY 816 AGAGCTGTAACTCACTGCGAAGCTTGTGAGCTTCACTCTGTAAGTCAAG-TGAGACAC 874  
Db 31818 AGAGCTGTAACTCACTGCGAAGCTTGTGAGCTTCACTCTGTAAGTCAAGTCAAG 31877  
QY 875 AAGCCACCAAGAAAGAAAGTCTGAGACACCTGTAATCTGAAAGAAAGAAAGTCAAG 934  
Db 31878 AAGCCACCAAGAAAGAAAGTCTGAGACACCTGTAATCTGAAAGAAAGAAAGTCAAG 31937  
QY 935 ACACCAATCTTTCAGAGCTG-TAACACTACCGCAAGGCTGTGGCTTCACTTGTGA 993  
Db 31938 ACACCAATCTTTCAGAGCTG-TAACACTACCGCAAGGCTGTGGCTTCACTTGTGA 31997  
QY 994 GTGAGCAAGCAAGAAAGTCTGAGACACCTGTAATCTGAAAGTCAAGTCAAG 1039  
Db 31998 GTGAGCAAGCAAGAAAGTCTGAGACACCTGTAATCTGAAAGTCAAGTCAAG 32043

RESULT 44  
US-09-949-016-16180  
; Sequence 16180, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16180  
; LENGTH: 157644  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(157644)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-16180

Query Match 17.5%; Score 349.2; DB 3; Length 157644;  
Best Local Similarity 82.7%; Pred. No. 2,4e-81;  
Matches 435; Conservative 0; Mismatches 88; Indels 3; Gaps 3;

QY	577	TAAGGGAGTAAAACTGGCCACCGAGCCAGAGTGGCAACCACCTGGGATCCCTTCCA	636
Db	31578	TAAGGAATATTAACGCGAGCAACTGAGCCAGACGACGACACTCCTGGGTACCTTGGG	31637
QY	637	CAGTGGGAAGCTTTGTTCTTTTGCTCTTCACATAATATCTGGCTGCTCATTCTTTG	696
Db	31638	TGTTGTGAAGCTTTGTTCTTTTCATTCTTGGCAATTAATCTGGTCTGCTCATTCTTTG	31697
QY	697	TGTTCCACATACCTTTATGAGCTGTGACATCTCATCTCGAGGGTGTGTGCTTCATTCCG	756
Db	31698	GGTTTGCTGTGCTTTATGAGCTGTGAACCTCATCTGGAAGTTGTCAACTTCACCTTCG	31757
QY	757	AAGTCAAC-AGACCACGAAACCCACTGGAGAAACAAGAACTCCGANTGTGCTTTA	815
Db	31758	AAGCAACAGAACCATGAACTCTGTGGCAGAAAGAAACAACCTTGACACGGCACTTGA	31817
QY	816	AGAGCTGTAAACCTCACTCGGGAGCTTGGACGTTCACTCTGTAAGTCAG-TGAGACAC	874
Db	31818	AGAGCTGTGACACTGACCGCGGAAGTCTGAGGCTTCACTCTGTAATTCAGCATTGAAC	31877
QY	875	AAACCCACCGAAGGAAGAAACTGTGGACACCTGTAATTTGGAAGAAACAACCTCGAG	934
Db	31878	GAACTCCACGAAAGGAAGATATCTCCGACACATCTGAACAGCTGGAAGAAACAACCTCG	31937
QY	935	ACACACCATCTTTGAGAGCTG-TAACACTCAACCGCAAGGCTCTGTGACTTCATTCTGAA	993
Db	31938	ACAACACATCTTTTAAGAACTGTTAACTCAACCGCGAGGGTCCGACGCTTCAATTCTTGA	31997
QY	994	GTCAGCAAGACCAAGAAACCCACCGGAAGAAACAATTCCAGACACA	1039
Db	31998	GTCCGCGAGACCAAGAAACCCACTTAAGTATTCATTTCTGGAACA	32043

```

RESULT 45
US-09-949-016-14659/C
; Sequence 14659, Application US//09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 14659
; LENGTH: 22888
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14659

```

Query Match	17.4%;	Score 348.6;	DB 3;	Length 22888;
Best Local Similarity	73.4%;	Pred. No. 1.5e-81;		
Matches 550; Conservative	0;	Mismatches 119;	Indels 80;	Gaps 5;

QY 371 GCAGCTGGGGGTGCTCTGTTTGAGAGGGGAGTGAAGTGAAGCACTGGGCTTCG8 430

Db 15993 GAGGACCACTGTCTTGCGCTCTCTGATCAAAATGAGAGTGAAGTGGCTGGGCTTCG8 15933

QY 431 TCAGGTGGGGACTTGGAGACCTTTTGTTGTCTAGCTGAAGATTCGTAATGCACCAATCAG 490

Db 15933 TCAGGTGAGGACTTGGAGACCTTTCTGCTCCTAGCTGAAGATTCGTAATGCACCAATCAG 15874

QY 491 CACTTGTGTCTAGCTTAAGATTTGTAATGCACCAATCAGCACTCTGTAA----- 542

Db	15873	CGCTCTGTGTCGTAGCAAAAGGTTTGTAAACGACCAATCAGACCTCTGTAAAAAAGACACC	15814
QY	543	-----ATGACCAATCAG-----	555
Db	15813	AATCAGACACTCTGTAAAAACGACCAATCAGACCTCTGTAAAAATGACCAATCAGCTCTCT	15754
QY	556	-----CAGGATGTGGCGGGGTCAATAAGGAGGT-----AAAAACT	592
Db	15753	GTAAAAATGACCAATCTGCATGTGGGTGGGGCCMAATTAAGGATTTCCCCCCCCACCTCC	15694
QY	593	GGCCACCCGAGCCAGACAGTGGCAACCCACTCTGGGTCCCTTCCACACTGTGAAG-CTTT	651
Db	15693	GCCCTTCCCAAGACAGTAGTACCACTGGCTCAGTGTCTTCCAGGCTGTGGAAGTGTAT	15634
QY	652	GTCTTTTGGCTCTTCCAAATTAATCTTGGCTGCTCATCTTTTGTGTCCACACTACCTT	711
Db	15633	TTTCTTTTCACTCGTCAACATTAATCTCGCTGTGTCTCACTCTTTGGGTTTGAATCACTTT	15574
QY	712	TATAGCTGTAACTCACTCACTCGACGAGGTCGTGGCTTCAATTCCTGAAGTCA-ACAGACCA	770
Db	15573	TATAGCTTAATTAACCTTCACTATATGAAGTCTCGAGCTTCACTCTCGAAGCAGCAGACCA	15514
QY	771	CGAACCACTGTGAAGGAACAAAGAACTCCCGATGTGCTGCTTTAAGACTGTAACTACCTC	830
Db	15513	TGAACCCACCAAGAGGAATTAACAATCCACGACGCCACCTTTAAGACTGTAGCACTC	15454
QY	831	ACTGGGAAGCTCTGACACTTCACTCCGGAAGTCAAGTGAAGCCACAAACCCACAGAAAGGA	890
Db	15453	ACTGGGAAGGCTGCGGGCTTCACTCCGTAAGTCAAGCAAGACCAACCAACCCACCGAAGGA	15394
QY	891	AGAAACTCTGACACACCTGAATATCTGAAGGAACAACTCCAGACACACATCTTTGAG	950
Db	15393	AGAAACTGCGGACACATCTGAACATCTGAAGGAACAACTCCGGAACACACATCTTTAAG	15334
QY	951	AGCTGTAACTCACTCACTCGGAAGGCTCTGTGGCTTCACTTTGAAGTCAAGCAAGCAAGAA	1010
Db	15333	AACGTAACTCACTCACTCGGTGAAGCCCGGCTTCACTTTAGAGTCAGACAGCAAGCAAGAA	15274
QY	1011	CCGACCGGAAGGAACAATTCGAGACA 1039	
Db	15273	CTGACGAGGAAGATCAATTTTGGACACA 15245	

Search completed: January 22, 2006, 21:31:03  
Job time : 275.224 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2006, 16:10:03 ; Search time 5419.54 Seconds  
(without alignments)  
17274.699 Million cell updates/sec

Title: US-09-728-552A-3\_COPY\_78595\_80595

Perfect score: 2001  
Sequence: 1 acccttattctgtatgagaaa.....aagctattctgcagaattc 2001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 100 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	* Match	Query Length	DB ID	Description
1	650.4	32.5	652	9	B72263
2	478.8	23.9	517	9	AQ251771 HS_3202_B
3	453	22.6	466	9	AQ737783 HS_2250_B
4	423	21.1	434	9	AQ593093 HS_5457_A
5	346	17.3	359	9	AQ822612 HS_5558_A
6	249	12.4	660	9	AQ353839
7	165	8.2	792	10	AG540684 Mus muscu
8	163.2	8.2	597	9	AZ706600 RPI-23-2
9	144	7.2	375	9	AZ065413 RPI-23-3
10	140.8	7.0	544	9	AQ463585 HS_5212_A
11	121.8	6.1	707	10	AG310514 Mus muscu
12	72.6	3.6	550	9	AZ879366 RPI-23-1
13	67	3.3	822	10	AG282114 Mus muscu
14	57.4	2.9	905	10	CNS00KXJ
15	53.8	2.7	734	9	AQ478276 RPI-11-2
16	53.2	2.7	996	10	CNS00RTH
17	51.8	2.6	997	10	CNS005TE
18	51.8	2.6	1200	10	CNS016CO
19	51	2.5	987	10	CNS014PO
20	50	2.5	1101	10	CNS0039G
21	48.8	2.4	377	9	AQ212826 HS_3118_B
22	48.8	2.4	1062	1	AJ925404



C 96 44 2.2 550 6 CA848973  
97 44 2.2 618 5 BQ632119  
C 98 44 2.2 699 2 BG180801  
C 99 44 2.2 742 5 BU852860  
100 44 2.2 775 5 BU626054

## ALIGNMENTS

RESULT 1  
B72263/c 652 bp DNA linear GSS 08-APR-1999  
DEFINITION RPI11-7022.TV RPI1-11 Homo sapiens genomic clone RPI1-11-7022,  
genomic survey sequence.  
ACCESSION B72263  
VERSION B72263.1 GI:2711414  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 652)  
Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
Venter,J.C.  
Use of BAC End Sequences for Sequence-Ready Map Building  
Unpublished (1997)  
Other\_GSSs: RPI11-7022.TP

TITLE  
JOURNAL  
COMMENT  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org

Clones are derived from the human BAC library RPI1-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers

FEATURES  
source  
1..652  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7502661"  
/db\_xref="taxon:9606"  
/clone="RPI1-11-7022"  
/sex="Male"  
/cell\_type="lymphocytes"  
/clone\_1ib="RPI1-11"  
/note="Vector: PBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPI11 Human Male BAC Library"

## ORIGIN

Query Match 32.5%; Score 650.4; DB 9; Length 652;  
Best Local Similarity 99.8%; Pred. No. 4.5e-157;  
Matches 651; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1070 GTGGAACTGACCTCTTATGTGTGTCAGCCTTTCTTCTTCCACAGAGCAGCTGTGTTTC 1129  
DB 652 GTGGAACTGACCTCTTATGTGTGTCAGCCTTTCTTCTTCCACAGAGCAGCTGTGTTTC 593  
QY 1130 TGGTCACTCTCCATAGGAACATCATGCTCTGATCTTCAGACCAACATCTGAGATGATAG 1189  
DB 592 TGGTCACTCTCCATAGGAACATCATGCTCTGATCTTCAGACCAACATCTGAGATGATAG 533  
QY 1190 TGGTCTTGACAGTCTTGAAGATTGTCTTACCGCTGATCTTCAAAAGCGGTGTGACACCGT 1249

DB 532 TGGTCTTGACAGTCTTGAAGATTGTCTTACCGCTGATCTTCAAAAGCGGTGTGACACCGT 473  
QY 1250 GAGAGAGAATAGAGAACTGGGCTCTTCAGGTAATCTTGCTTTTTCACAAGCCCCCTA 1309  
DB 472 GAGAGAGAATAGAGAACTGGGCTCTTCAGGTAATCTTGCTTTTTCACAAGCCCCCTA 413

QY 1310 ATTTTACTGCATTAATTTATTTTGAATTCACGTAAATTTCTACAAATTTTCCATAAGTCAT 1369  
DB 412 ATTTTACTGCATTAATTTATTTTGAATTCACGTAAATTTCTACAAATTTTCCATAAGTCAT 353

QY 1370 CTACACACATACCTCTTCATGCAACACTTGCTTGTCTTAATACATCTATTAAGAG 1429  
DB 352 CTACACACATACCTCTTCATGCAACACTTGCTTGTCTTAATACATCTATTAAGAG 293

QY 1430 CTGTCTCTTCTTAAGGCTTAATGTTTATATGACATAAGGCTCTTGCTTACATATAAG 1489  
DB 292 CTGTCTCTTCTTAAGGCTTAATGTTTATATGACATAAGGCTCTTGCTTACATATAAG 233

QY 1490 GGGTATGAGCAATGTATACAGAAAGCTTTTTCACAGAGCTCATATGTAAAGATTC 1549  
DB 232 GGGTATGAGCAATGTATACAGAAAGCTTTTTCACAGAGCTCATATGTAAAGATTC 173

QY 1550 ATTAAATGGCTGAANTAGACTGATCTGTCATTTCTGCTCACTTATCATTAAGAGT 1609  
DB 172 ATTAAATGGCTGAANTAGACTGATCTGTCATTTCTGCTCACTTATCATTAAGAGT 113

QY 1610 CATTAGCTTAGGACAAACAACTACATCTATGTAATTAAGAGACAGCTGTTTGCTC 1669  
DB 112 CATTAGCTTAGGACAAACAACTACATCTATGTAATTAAGAGACAGCTGTTTGCTC 53

QY 1670 AATATTAATAATTAAGAAAGAAACCATGTGAAGTCAAAATATTTGTTAAT 1721  
DB 52 AATATTAATAATTAAGAAAGAAACCATGTGAAGTCAAAATATTTGTTAAT 1

RESULT 2  
AQ251771/c 517 bp DNA linear GSS 10-OCT-1998  
LOCUS HS 3202 B1 G12 T7 C1F Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone Plate=3202 Col=23 Row=N, genomic survey  
sequence.

ACCESSION AQ251771  
VERSION AQ251771.1 GI:3722344  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 517)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3202 Row: N Column: 23  
Class: BAC ends  
High quality sequence scop: 517.

FEATURES  
source  
1..517  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"



**AUTHORS** Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

**JOURNAL PUBLISHED** Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

**COMMENT** 10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
Plate: 1033 row: K column: 20  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 434.

**FEATURES**  
Location/Qualifiers  
source  
1..434  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=1033 Col=20 Row=K"  
/sex="male"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

**ORIGIN**

Query Match 21.1%; Score 423; DB 9; Length 434;  
Best Local Similarity 98.4%; Pred. No. 3.7e-98;  
Matches 426; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

1110 CAGAGTCAGCTGCTGTTTCTGCTGACTCTCTCATAGGAACATGCTGATCTCTCAGA 1169  
433 CAGAACTCAGCTGTGTTGTTGCTGACTCTCCATAGGAACATGCTGATCTCTCAGA 374  
1170 CCACCATCTGAGTAGTAGTGTCTCTGACAGTCTAGAAAGTTGTCTACCGCTGATCTC 1229  
373 CCACCATCTGAGTAGTAGTGTCTCTGACAGTCTAGAAAGTTGTCTACCGCTGATCTC 314  
1230 CAAGAGGTGACACACCGTGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAATCTT 1289  
313 CAAGAGGTGACACACCGTGAGAGAAATGAGAAAGCTGGGCTCTTCAGGTAATCTT 254  
1290 GCTTTTTCACAAAGCCCTTAATTTTACTGCATATTAATTTTGAATTCATGATATTTCT 1349  
253 GCTTTTTCACAAAGCCCTTAATTTTACTGCATATTAATTTTGAATTCATGATATTTCT 194  
1350 ACAATTTTCCATAGTCATCTACACACATATACCTCTCATGCAACACTGGCTGGCTGA 1409  
193 ACAATTTTCCATAGTCATCTACACACATATACCTCTCATGCAACACTGGCTGGCTGA 134  
1410 ATTCATATCTATATAGAGAGCTGTCTTTAAGCCGTAATGTTTATATAGCACTAAGC 1469  
133 ATTCATATCTATATAGAGAGCTGTCTTTAAGCCGTAATGTTTATATAGCACTAAGC 74  
1470 TCTTGCTTACATATAAAGGGGTATTGACCAATGTATACAGAAGCTTTTCTCCACAG 1529  
73 TCTTGCTTACATATAAAGGGGTATTGACCAATGTATACAGAAGCTTTTCTCCACAG 14  
1530 GTCTCATATGTAA 1542  
|||||||

**Db** 13 GTCTCATATAAAAA 1

**RESULT 5**  
LOCUS A0822612/c 359 bp DNA linear GSS 26-AUG-1999  
DEFINITION HS\_5558\_A1\_F10\_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1134 Col=19 Row=K, genomic survey sequence.  
ACCESSION A0822612  
VERSION A0822612.1 GI:5785005  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 359)  
REFERENCE 1 (bases 1 to 359)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

**JOURNAL PUBLISHED** Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

**COMMENT** 10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
Plate: 1134 row: K column: 19  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 359.

**FEATURES**  
Location/Qualifiers  
source  
1..359  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=1134 Col=19 Row=K"  
/sex="male"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

**ORIGIN**

Query Match 17.3%; Score 346; DB 9; Length 359;  
Best Local Similarity 100.0%; Pred. No. 3.2e-78;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1413 CATATCTATATAGAGAGCTGCTCTTAAGCGTAATGTTTATATAGCACTAAGGCTCT 1472  
359 CATATCTATATAGAGAGCTGCTCTTAAGCGTAATGTTTATATAGCACTAAGGCTCT 300  
1473 TGGCTTACATATAAAGGGGTATTGAGCAATGTGATACAGAAGCTTTTCTCCACAGTC 1532  
299 TGGCTTACATATAAAGGGGTATTGAGCAATGTGATACAGAAGCTTTTCTCCACAGTC 240  
1533 TCATATGTAAAGAAATTCATTAGATTGGCTGAATAGACTGATCTGTCTCTGCTC 1592  
239 TCATATGTAAAGAAATTCATTAGATTGGCTGAATAGACTGATCTGTCTCTGCTC 180  
1593 ACTTATCATTAAGAAAGTCAATTAGCTAAGGAACAAAACTAACAATCTATGTAAATTAGAAGA 1652  
|||||||

|||||  
Db 179 ACTTACATAGAGAGATCTAGCTAGAGAAACAACTACATCTATGTATATAGAGA 120  
|||  
QY 1653 ACAAGCTGCTTTTCTCAATATATAAATAAGAAAAAGAACATGTGAAGTCAAAATAT 1712  
|||  
Db 119 ACAAGCTGCTTTTCTCAATATATAAATAAGAAAAAGAACATGTGAAGTCAAAATAT 60  
|||  
QY 1713 TTGTTTAAATCAGGTCATGTGAGATCTTATTTAAAGTATTTGAATTC 1758  
|||  
Db 59 TTGTTTAAATCAGGTCATGTGAGATCTTATTTAAAGTATTTGAATTC 14  
|||  
RESULT 6  
AO353839 660 bp DNA linear GSS 07-MAY-1999  
LOCUS RPCI11-109P10.TV RPCI-11 Homo sapiens genomic clone RPCI-11-109P10,  
DEFINITION genomic survey sequence.  
ACCESSION AO353839  
VERSION AO353839.1 GI:4176797  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
1 (bases 1 to 660)  
Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and  
Venter,J.C.  
Use of BAC End Sequences from library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
Other\_GSSs: RPCI11-109P10.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@igr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: SP6  
Class: BAC ends.  
FEATURES  
source Location/Qualifiers  
1..660  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7541841"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-109P10"  
/sex="male"  
/cell\_type="Lymphocytes"  
/clone\_lib="RPCI-11"  
/note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPCI11 Human Male BAC Library"  
ORIGIN  
Query Match 12.4%; Score 249; DB 9; Length 660;  
Best Local Similarity 100.0%; Pred. No. 5.5e-53;  
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
|||  
Db 1753 GAATTCCTTATGATGAGACTATCTGACTCAAGTGAGAGCTTTTGGCCTG 1812  
|||  
QY 5 GAATTCCTTATGATGAGACTATCTGACTCAAGTGAGAGCTTTTGGCCTG 64  
|||  
Db 1813 TGGTCCCTACGTAGAAAGAGGCTTTGTCTATAAAGTCTTATATGTACAGGTCCAAAGTT 1872  
|||  
QY 65 TGGTCCCTACGTAGAAAGAGGCTTTGTCTATAAAGTCTTATATGTACAGGTCCAAAGTT 124  
|||

|||  
QY 1873 AAGTCCCAAGCTTGCTCTTAAAGCATCTGATTTGTTTATAGACTTGAACCTG 1932  
|||  
Db 125 AAGTCCCAAGCTTGCTCTTAAAGCATCTGATTTGTTTATAGACTTGAACCTG 184  
|||  
QY 1933 AAGGGAATTAACAAATCCCTCTGGAGAACTTCTCTCATCCTTGGTGAAGTCAATTCG 1992  
|||  
Db 185 AAGGGAATTAACAAATCCCTCTGGAGAACTTCTCTCATCCTTGGTGAAGTCAATTCG 244  
|||  
QY 1993 CCAGATTC 2001  
|||  
Db 245 CCAGATTC 253  
|||  
RESULT 7  
AG540684/c 792 bp DNA linear GSS 23-DEC-2004  
LOCUS AG540684  
DEFINITION Mus musculus molossinus DNA, clone:MSWg01-453L21.TV, genomic survey  
sequence.  
ACCESSION AG540684  
VERSION AG540684.1 GI:48301098  
KEYWORDS GSS.  
SOURCE Mus musculus molossinus (Japanese wild mouse)  
ORGANISM Mus musculus molossinus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1  
Abe,K., Noguchi,H., Tagawa,K., Yuzuriba,M., Toyoda,A., Kojima,T.,  
Ezawa,K., Satou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and  
Shiroishi,T.  
Contribution of Asian mouse subspecies Mus musculus molossinus to  
genomic constitution of strain C57BL/6J, as defined by BAC-end  
sequence-SNP analysis  
Genome Res. 14 (12), 2439-2447 (2004)  
1574823  
2 (bases 1 to 792)  
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
Direct Submission  
Submitted (17-NOV-2003) Maabihira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Shuho-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the mouse BAC library MSWg01. For BAC  
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyada, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : TV  
LIBRARY  
Vector : pBACE3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.  
FEATURES  
source Location/Qualifiers  
1..792  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSWg01-453L21.TV"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSWg01 Mouse Male BAC Library"  
ORIGIN  
Query Match 8.2%; Score 165; DB 10; Length 792;  
Best Local Similarity 63.8%; Pred. No. 3.5e-31;  
Matches 340; Conservative 0; Mismatches 170; Indels 23; Gaps 5;  
|||  
QY 1044 TAGTCTCAGCAGCTGACAGCTGAGAACTGACTCTTATGTGTGTCAGGCTTTC 1103  
|||

Db 590 TACTATGTGTGATACAGTCTCAGAGAGAGAACCCACCTTCTGTGTGGCCAGCTTCC 531  
QY 1104 TTCCTTCAAGAGTACAGTGTGTTTCTGTGACTCTCCATAGAAATAGT-CCGTAAT 1162  
Db 530 CTCGAAATGGCCCTCTTCTTATGTGATGAGTCCCTCAGAGATGACATCGTAAT 471  
QY 1163 CCTCAGACCAACATCTGAGTAGTGTCTCTGACAGTCCAGAGATGTCTACCGCT 1222  
Db 470 CCCCACCAACCTCCCTGAGTACTAGCATTTCTGTGAGTTACACAGCCGCCACCCCG 411  
QY 1223 GGATCTCCAAAGCGTGTGACACCCGTGAGAGAGAAATGAGAAAGCTGGCTCTTCAAGT 1282  
Db 410 GGAGCCCAAGTCTCTGTGTGAGACTATGATAGAGATGAGAGAGGTCTGAGCTTTTCAAGT 351  
QY 1283 AAA-----TCTGCTTTTTCACAGCCCTTAACTTTTACTGATTAATATT 1329  
Db 350 AAACCTTTCCTCTTGTGTGTGTTTAAACAGCCCATAGTTTCACTGATAGTGTCT 291  
QY 1330 TGAATTCAGTAAATTTTCTACAAATTTCCCATAGTCAATCTACACAAATACCTCTCA 1389  
Db 290 TGATTCACATGAAATTTGGGCAATTTCCCTGTAAGTCAATCAAAATGTTCCCTTAA 231  
QY 1390 TGCACACCTTGGCTTGTCTAATACAT-----ATCTAATATGAGAGCTGTCTTCTTAA 1443  
Db 230 TGTGGCCCTCAGCTATGCGGCCACGTTATCTGTCTAATTAAGAGAGCTGTCTTCTTGA 171  
QY 1444 CGTAATGTTTATGATGACATAAGGCTCTGCTTACATTAAGAGGATATGAGCAAT 1503  
Db 170 TATTAATGTTTATGATGATTAAGGCTTGGCTTACATAT-CAAGGACATTTGACAAAC 112  
QY 1504 GTGATACAGAGTCTTCTTCCACAGTCTCATATGTAAGATTCATTAGAT 1556  
Db 111 GGTATACAGAGCTTTTGCCTGAC--GACTCGTGTGTAAGATTCGGCGAT 61

RESULT 8  
AZ706600 597 bp DNA linear GSS 24-JAN-2001  
LOCUS RPCI-23-209G10.TJ RPCI-23 Mus musculus genomic clone  
DEFINITION RPCI-23-209G10, genomic survey sequence.  
ACCESSION AZ706600  
VERSION AZ706600.1 GI:12434225  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 597)  
Zhao,S., Nieman,W., Feldblyum,T., Malek,J., Shatsman,S.,  
Akimov,B., Levins,M., Megam,S., Tsegeye,G., Geer,K., Krol,M., de  
Jong,P. and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other GSSs: RPCI-23-209G10.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 209 row: G column: 10  
Seq primer: Sp6  
Class: BAC ends.  
Location/Qualifiers  
1..597

FEATURES  
Source

/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-209G10"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:  
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
ORIGIN

Query Match 8.2%; Score 163.2; DB 9; Length 597;  
Best Local Similarity 65.5%; Pred. No. 9,4e-31;  
Matches 325; Conservative 0; Mismatches 148; Indels 23; Gaps 5;  
QY 1044 TGTCTCAGAGGAGTGCAGGCTCAGAGTGAAGTGAACCTGACCTTATGTTGTCAGCTTTC 1103  
Db 494 TACTATGTGTGATACAGTCTCAGAGAGAGAGCCACATTTCTGTGTGGCCAGCTTCC 435  
QY 1104 TTCCTTCAAGATCAGCTGTGTTTC--TGCTGACTCTCCATAGGACATCAGTCTGAA 1161  
Db 434 CCTGAAATGAGCCCTCTTGGCTTTCTATGTGATAGCCCTCAGAAATGCAATGACTGAA 375  
QY 1162 TCTTCAAGACCAATCTGAGATGTAAGTCTCTCTGACAGTCTTGAAGTTCTTACCGC 1221  
Db 374 TCCCAAAACACCTCTGAAATGATGATGAT-CTGTGATTCACAGAGCCGCCACCCCC 316  
QY 1222 TGGATCTCCAAAGCGTGTGACACACCGTGAAGAGAAATGAGAAAGCTGGGCTTTCAG 1281  
Db 315 GGAGCCCAAGTTCCTGTGTGATGAGAGAGAAATGAGAGGCTGACCTTTTTCAGA 256  
QY 1282 TAAA-----TCTGCTTTTTCACAGCCCTTAACTTTTACTGCAATTAAT 1328  
Db 255 TAAACTTGGCTCTTGTGTTTCTTTTAAACAGCCCATATTTCACTGCAATGTTGCT 196  
QY 1329 TGAATTCAGTAAATTTTCTCAAAATTTCCATTAAGTCAATCAACATACCTCTTC 1388  
Db 195 TGAATTCAGTAAATTTTGGGCAATTTCCCTGAATGATTCACAAATAGTCTCCCTCG 136  
QY 1389 ATGCAACACTTGGCTTGTGATATACATATCT-----ATTATAGAGCTGTGCTCTTAA 1443  
Db 135 ATGTGCTCCTCAGCTATGCGCCAGATATCTATTAATTAAGAGCTGTGCTTGAAG 76  
QY 1444 CGTAATGTTTATATGACATAAGGCTCTTGGCTTACATATAAGGGTATTTAGCAAT 1503  
Db 75 TATTAATGTTTATGTGATTAAGGCTCTTGGCTTACATATCA--GGGCGATGAGCAAC 18  
QY 1504 GTGATACAGAGTCTT 1519  
Db 17 GAGATACAGAGCTT 2  
RESULT 9  
AZ065413 375 bp DNA linear GSS 30-MAR-2000  
LOCUS RPCI-23-391K3.TVB RPCI-23 Mus musculus genomic clone RPCI-23-391K3,  
DEFINITION genomic survey sequence.  
ACCESSION AZ065413  
VERSION AZ065413.1 GI:7356665  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 375)  
Zhao,S., Nieman,W., Feldblyum,T., Malek,J., Shatsman,S.,

TITLE  
JOURNAL  
COMMENT

Akinret, B., Levins, M., Mcgann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C. M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other GSSs: RPCI-23-391K3.TU  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/orderingframe.htm>) or from Resear ch Genetics ([inforesgen.com](http://inforesgen.com)). BAC end page: [http://www.tigr.org/cdb/bac/ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac/ends/mouse/bac_end_intro.html)  
Plate: 391 row: K column: 3  
Seq primer: T7  
Class: BAC ends.

# FEATURES

source  
1. .375  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-391K3"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

## ORIGIN

Query Match  
Best local Similarity 69.4%; Pred. No. 7.9e-26;  
Matches 261; Conservative 0; Mismatches 95; Indels 20; Gaps 4;

1161 ATCTCAGACACCATCTGAGTAGTGTCTCTGACAGTCTTGAAGTGTCTACG 1220  
|||||  
375 ATCCCAACACCTCTGAGTAGTGTCTCTGACAGTCTTGAAGTGTCTACG 317  
|||||  
1221 CTGGATCTCCAAAGCGTGTGACACACCGTGAAGAGAGAAATGAGAAAGCTGGCTTCAG 1280  
|||||  
316 CGGAGACCCCAAGTCTGTGTGTAAGTGAAGAGAGAAATGAGAAAGGCTTCAG 257  
|||||  
1281 GTAAA-----TCTTCTTTTTCACAAGCCCCCTAATTTTATGCAATTAAT 1327  
|||||  
256 ATAACTTGGCTCTTGTCTTTTGAACAGCCATATGTCATGCAATGTCG 197  
|||||  
1328 TTGAATTCATGTAATTTCTAACAATTTTCCATTAAGTATGATACACAAATACCTCT 1387  
|||||  
196 TTGAATTCATGTAATTTCTAACAATTTTCCATTAAGTATGATACACAAATAGTTCCTCT 137  
|||||  
1388 CATTGCAACCTGGCTTGTGTAATATCT-----ATTATGAGAGCTGTGCTTTTAA 1442  
|||||  
136 GATGTTGCTCTGATGCGGACGATCTGTCTATTAAGAGCTGTGCTTTGAA 77  
|||||  
1443 GCCTAATGTTTATATGCACTAAGGCTTGTGCTTACATATATAAAGGGTATTTGACAA 1502  
|||||  
76 GTATAATGTTTATATGCACTAAGGCTTGTGCTTACATAT-CAAGGGGCAATTGACAA 18  
|||||  
1503 TGTGATACAGAGCTCT 1518  
|||||  
17 CGAGATACAGAAAGCT 2

RESULT 10  
A0463585  
LOCUS

DEFINITION  
HS\_5212\_A1\_E01\_SPEE RPCI-11 Human Male BAC library Homo sapiens  
genomic clone Plate=78 Col=1 Row=I, genomic survey sequence.

ACCESSION  
A0463585  
VERSION  
A0463585.1 GI:4640680

KEYWORDS  
GSS.  
SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE  
1 (bases 1 to 544)  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
Hood, L.

AUTHORS  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
Hood, L.

TITLE  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

PUBMED  
10449764  
COMMENT  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.bufileo.edu/ordering\\_bac.htm](http://bacpac.med.bufileo.edu/ordering_bac.htm)) or from Resear ch Genetics ([inforesgen.com](http://inforesgen.com)). BAC end web Server: <http://www.htc.washington.edu>  
Plate: 788 row: I column: 1  
Seq primer: SP6  
Class: BAC ends

High quality sequence stop: 544.  
Location/Qualifiers  
1. .544  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=78 Col=1 Row=I"  
/sex="male"  
/clone\_lib="RPCI-11 Human Male BAC library"  
/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

## FEATURES

source  
1. .544  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=78 Col=1 Row=I"  
/sex="male"  
/clone\_lib="RPCI-11 Human Male BAC library"  
/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

## ORIGIN

Query Match  
Best local Similarity 74.9%; Pred. No. 5.9e-25;  
Matches 191; Conservative 0; Mismatches 58; Indels 6; Gaps 1;

1751 TTGAATCTTTATGATGAGAGTATCTTGTACTC-----AAGTGAAGAGTGTGACCTTT 1804  
|||||  
129 TTGAATCTTTATGAGAGTATCTTGTACTC-----AAGTGAAGAGTGTGACCTTT 188  
|||||  
1805 TTGGCTGTGTGCTCTGATGAGAGAGGCTTTGTCTAATAAGTCTTATATGATACAGT 1864  
|||||  
189 CTGGCTGTGTGCTCTGATGAGAGAGGCTTTGTCTAATAAGTCTTATATGATACAGT 248  
|||||  
1865 GCAAGTTAAGGCCCAAGCTTCTTAAAGCATGAGATTTTGTATGACTTTTA 1924  
|||||  
249 GCAACATTAATGAGCTTAACCTTGTCTTAAAGCATGAGATTTTGTATGACTTTTA 308  
|||||  
1925 GTGAAGTGAAGGAGATTAACAATCTCTGTGAGAGACTTCTCTCATCTTGTGGAAG 1984  
|||||  
309 CTGAAGCTGAAGGAGATTAACAATCTCTGTGAGAGACTTCTCTCATCTTGTGGAAG 368

```

QY      1985 TCATTCTGCCAGAAAT 1999
      |||||
Db      369 TCATTCTCCCNCACT 383

RESULT 11
AG310514      707 bp  DNA      linear  GSS 18-DEC-2004
LOCUS      Mus musculus molossinus DNA, clone:MSMg01-092P04.TU, genomic survey
DEFINITION
ACCESSION      AG310514
VERSION      AG310514.1  GI:47883468
KEYWORDS      GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
ORGANISM      Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      Abe, K., Noguchi, H., Tagawa, K., Yuzurika, M., Toyoda, A., Kojima, T.,
      Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaiki, K. and
      Shiroishi, T.
TITLE      Contribution of Asian mouse subspecies Mus musculus molossinus to
      genomic constitution of strain C57BL/6J, as defined by BAC-end
      sequence-SNP analysis
JOURNAL      Genome Res. 14 (12), 2439-2447 (2004)
PUBMED      15574823
AUTHORS      Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
TITLE      Direct Submission
COMMENT      Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
      and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
      1-7-22 Suenhiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
      (E-mail:hattori@gsc.riken.jp, URL://http://hgp.gsc.riken.go.jp/,
      Tel:81-45-503-9111, Fax:81-45-503-9170)
      Clones are derived from the mouse BAC library MSMg01. For BAC
      library availability, please contact Kun'ya Abe (abe@erc.riken.jp).
      Tsukuba Institute, Bio Resource Center,
      The Institute of Physical and Chemical Research (RIKEN) 3-1-1
      Koyada, Tsukuba, 305-0074 Japan
      phone: 81-298-36-9189, fax: 81-298-36-9199
      e-mail: abe@erc.riken.jp
      PRIMERS
      Sequencing : TU
      LIBRARY      : pBACe3.6
      R.Site 1 : EcoRI
      R.Site 2 : EcoRI.
      Location/Qualifiers
        1. 707
          /organism="Mus musculus molossinus"
          /mol_type="genomic DNA"
          /sub_species="molossinus"
          /db_xref="taxon:57486"
          /clone="MSMg01-092P04.TU"
          /sex="male"
          /tissue_type="mixture of kidney and spleen"
          /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match      6.1%; Score 121.8; DB 10; Length 707;
Best Local Similarity 63.2%; Pred. No. 5.4e-20;
Matches 265; Conservative 0; Mismatches 127; Indels 27; Gaps 4;

QY      1544 GAATTCATAGATTTGGTGAATAGACTGATCTGTCATTCTCTGCTCATATCAATA 1603
      |||||
Db      95 GAATTCACGTGGTGGTGAATAAGACTGATCTGTCATTTCGCTCATCTCCAGCAA 154

QY      1604 GGAAGTCATTA-----GCTAAGAAACAAACTCAATCTATGTATTAAGAAAGCAAGCT 1659
      |||||
Db      155 GGAAGTCATTAACCAAGCAACCGAAGGTAAGTGTATGCAATTAAGAAAGAGTT 214

QY      1660 GGTTCCTCATATATAAATAAGAAAGAAACCAATGTAAGTCAAAATATTTGTTTA 1719

```

```

Db      215 GCGTTTGCTCAATACAGAA-----ACAAAGAAACCATGCAAGAGACAGTCTTTCTTTA 269
      |||||
QY      1720 ATCAGGATCTGAGATCATTAATAAAGATTTGAATCTTTATGATGAGAACTAT----- 1775
      |||||
Db      270 ATTAGAGACTGAGATTCTATTAAGACACTTATGATATATTAATTTTCTTTTANGG 329

QY      1776 -----CTTACTCAAGTGACAGTGTGTGAGCTTTTGGCTGTGCTCCACTAGT 1824
      |||||
Db      330 TGGCAGCTATCTCTTGACTCAAGTGGAGAGTGTGCTCTCT---GGCTTTGGACCTATGC 386

QY      1825 AGAAGAGAGCTTTTGTCTATAAGCTTTATATGTAACAGGTGCCAAGTTAATGCCAAGC 1884
      |||||
Db      387 AGAAGAGAGCTTTTGTCTATAAGCTTTAAGCTCAAAATGTGCTCGGTGGAATCAAGGTTCCAGC 446

QY      1885 TTGCTCTTAAGAACATCTAGTGAATTTTGTGTTTATGACTTTTATGAACTGAGAGGAATTA 1943
      |||||
Db      447 TTGCTCTTAACATTTCTGGATGGTTTTTATATCCAGAACTACTCTGTTTCAAGGCAATTA 505

RESULT 12
AZ879366      550 bp  DNA      linear  GSS 05-MAR-2001
LOCUS      RPCI-23-185B21.TU RPCI-23 Mus musculus genomic clone
DEFINITION
ACCESSION      AZ879366
VERSION      AZ879366.1  GI:13197432
KEYWORDS      GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatman, S.,
      Akiniec, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de
      Jong, P. and Fraser, C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL      Unpublished (1999)
COMMENT      Other GSSs: RPCI-23-185B21.TV
      Contact: Shaying Zhao
      Department of Eukaryotic Genomics
      The Institute for Genomic Research
      9712 Medical Center Dr., Rockville, MD 20850, USA
      Tel: 301 838 0200
      Fax: 301 838 0208
      Email: szhao@tigr.org
      Clones are derived from the mouse BAC library RPCI-23. For BAC
      library availability, please contact Pieter de Jong
      (pdejong@mail.cho.org). Clones may be purchased from BACPAC
      Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
      page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html
      Plate: 185 Row: B Column: 21
      Seq primer: SPE
      Class: BAC ends.
      Location/Qualifiers
        1. 550
          /organism="Mus musculus"
          /mol_type="genomic DNA"
          /strain="C57BL/6J"
          /db_xref="taxon:10090"
          /clone="RPCI-23-185B21"
          /sex="Female"
          /lab_host="DH10B"
          /clone_lib="RPCI-23"
          /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1;
      EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
      brain genomic DNA was isolated and partially digested
      with a combination of EcoRI and EcoRI Methylase. Size
      selected DNA was cloned into the pBACe3.6 vector at the
      EcoRI sites. The ligation products were transformed into
      DH10B electrocompetent cells (BRL life technologies)."
ORIGIN

```





Query Match	Similarity	Score	DB	Length	905;
Best Local	Similarity	36.5%	Pred. No.	0.0029;	
Matches	105;	Conservative	51;	Mismatches	132;
				Indels	0;
				Gaps	0;
QY	1493	TATTGACCAATGTCATGATACGAAAGTCTTTTCCTCCACAGGCTCTCATATGTAAAGAAATTCATT	1552		
Db	493	TNNTTGNCBMTWTMTTWTTTTAMTTTGTKTATTAATAATGTWGGKKKKRRMTTKTKDKTK	552		
QY	1553	AGATTGCGTGAATGATGATGATCTGTCATTTCTGCTCATTTATCTAATGAAGATCAT	1612		
Db	553	AAAATTATWAKAWMAAAATATKTKTWTDTTKWMTATKAAAAAAHAAWATAWATAWATA	612		
QY	1613	TAGCTAAGCAACAAAACCTACATCTATGTATTTAGAAACAAGCTGTTTGTCTCAAT	1672		
Db	613	TATATTAATAAAAAATTAATTAATAAAAAATTAATAAAATTAATTTATTAATAWATA	672		
QY	1673	ATAAAAATTAAGAAAAAGAAACCATGTCGAAGTCAAAATATTTGTTTAATCAGGTCAATGA	1732		
Db	673	AAAAAATAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAAAWMAATTTATTTATTA	732		
QY	1733	GAATCTATTAATAAAAGTATTTGAATTCCTTTAGCATAGACATCATCTGA	1780		
Db	733	WMAATTAATAAAWMAAAWMAAAWMAATTAATAWMAATTAATAWMAATTAATAWMAA	780		
RESULT 15					
LOCUS	ACQ478276/C	734 bp	DNA	linear	GSS 23-APR-1998
DEFINITION	RPCT-11-254C3. TJ RPCT-11 Homo sapiens genomic clone RPCT-11-254C3,				
ACCESSION	ACQ478276				
KEYWORDS	ACQ478276.1 GI:4660395				
SOURCE	GSS.				
ORGANISM	Homo sapiens (human)				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 734)				
JOURNAL	Zhao, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P. and Venter, J.C.				
COMMENT	Use of BAC End Sequences from Library RPCT-11 for Sequence-Ready Map Building				
	Unpublished (1997)				
	Other GSSs: RPCT-11-254C3, TV				
	Contact: Shaying Zhao, William Niernan, Mark Adams				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: hbe@tigr.org				
	Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong				
	(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Gene cs (info@resgen.com). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.				
	Seq primer: SP6				
	Class: BAC ends.				
FEATURES					
source	Location/Qualifiers				
	1..734				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="GDB:7597202"				
	/db_xref="taxon:9606"				
	/clone="RPCT-11-254C3"				
	/sex="Male"				
	/cell_type="Tlymphocytes"				
	/clone_1lb="RPCT-11"				
	/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCT11 Human Male BAC Library"				

ORIGIN	
Query Match	2.7%; Score 53.8; DB 9; Length 734;
Best Local Similarity	62.0%; Pred. No. 0.023;
Matches	85; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
Oy	4 TTATTTGTATGAGAAAAAGAAAAGACGTGAAATAATTATTAACCTGCATATAGTTC A 63
Dd	567 TTTCTTTGTCTGATTTAACAAGAATTGCTAATTAATTCACGCAATGATTTAGTTGTG 508
Oy	64 GGACCATGATTTGCACTGACAGAACTCAATTCAAACCAACGTACTCAAAAGAAAT 123
Dd	507 GATATGTTAATTGCACTGACAGAAACCCCTACTCAAACCTGGGTTAACCAAAATCGGAAT 448
Oy	124 ATATTGCTCATGTATAC 140
Dd	447 TTATTGGTTCATAAACC 431
RESULT 16	
CNS00FUH	996 bp DNA linear GSS 03-JUN-1999
Locus	
LOCUS	
DEFINITION	Drosophila melanogaster genome survey sequence YET3 end of BAC: BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL071063.1 GI:4951105
VERSION	GSS.
KEYWORDS	Drosophila melanogaster (fruit fly)
SOURCE	Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 996)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	location/Qualifiers
SOURCE	1..996
ORIGIN	/organism="Drosophila melanogaster"
	/mol_type="genomic DNA"
	/db_xref="taxon:7227"
	/clone="BACR31021"
	/clone_lib="RPCI-98"
	/note="end : YET3"
Query Match	2.7%; Score 53.2; DB 10; Length 996;
Best Local Similarity	45.1%; Pred. No. 0.036;
Matches	74; Conservative 28; Mismatches 62; Indels 0; Gaps 0;
Oy	1600 ATAAGCAGCTTAGCTTAGGAACAAACTACATCTATGTAATTAGAGAACAGCT 1659
Dd	736 ATAWAAATTAAMWTADATTWATMAAAAAAAAAAAAAAAAAMATTTWTTATATMAAMWAAT 795
Oy	1660 GGTITTGCTCAATATMAAAATMAAGAAAAAGAACCAAGTGAAGTCAAAATATTGTTTA 1719



Db		1143	TWMTWMATTTTATTATTTTA	1161
RESULT 19				
CNS014PQ				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
source				
ORIGIN				

```

CNS014PQ          987 bp      DNA           linear    GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SpE end of BAC
BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL104456
ALI04456.1       GI:5616067
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 987)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CPHF (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pbeloBAC11.

Location/Qualifiers
1..987
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN12P22"
/clone_lib="DrosBAC"
/plasmid="pbeloBAC11"
/note="end : SP6"

Query Match          2.5%; Score 51; DB 10; Length 987;
Best Local Similarity 30.5%; Pred. No. 0.13;
Matches 86; Conservative 68; Mismatches 128; Indels 0; Gaps 0;
```

QY	1440	TAAAGCTAATGCTTTTATATGCACTGAAGGCTCTTGCTTACATATAAAGGGATTCGC	14399
DB	687	TAAATATATWAMWTNTTWTTTTTTTTTTTTTTTAAAAAAWWAAAAMATWTTAWM	746
QY	1500	CAATGTGATACGAAGTCTTTTCTCCACAGGCTCATATGTAAAGATTGATGATGC	15599
DB	747	ATWTTTAAAGAAATTTATTAATTAATTAATTTTWTWTTATTAATTAATTAATTAATTA	806
QY	1560	CTGAATAGACTGATCTGCTCATCTCTGCTCACTTACATTAAGAAGTCATTAGCTAA	16199
DB	807	TTWATWMAATTAATATWMTWTANAKMTTWTTTTTTTTTTWTAATAATTTTWWAMAAA	866
QY	1620	GGAACAATAAATCTGCAATCTAIGTAATTAGAACACAGCTGGTTTGGCTCAATATAAA	16799
DB	867	TAMBRMTWMAAATAWMAAATWMAAATWMAAAMWMDRAAMDMMWMTAAAWTTAA	926
QY	1680	TAAAGAAAAGAACATGTGAAGTCAAATATTTGTTTAAT	1721
DB	927	AA	968

RESULT 20				
CNS0039G/c				
LOCUS				
DEFINITION				

Drosophila melanogaster genome survey sequence TE73 end of BAC #  
BACH08K10 of RPCI-98 library from Drosophila melanogaster (fruit

ACCESSION	fly), genomic survey sequence.
VERSION	AL063921
KEYWORDS	AL063921.1 GI:4941778 GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Empidoidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr )
REFERENCE	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the Isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACR08K10" /clone_lib="RPc1-98" /note="end : TET3"
ORIGIN	
Query Match	2.5%; Score 50; DB 10; Length 1101;
Best Local Similarity	19.4%; Pred. No. 0.25;
Matches	Conservative 232; Mismatches 302; Indels 3; Gaps 1
Db	4 TTTATTGTATGAGAAAAAAGAGAGTGAATAATATTAACGTGCATATAGTCA 63  :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : 901 DDTGGTKDDDKMDKMDWDKAKGTGDATWMAATDMWMGMADDMWTWDAADDWNA 842
Oy	64 GGACCATGATTCAGTAGCAGAAACTCAATTCAAACCAGCTAAGTCAAAAGAANAAT 123   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : Db DRMDMAMKWMDWAMWGARTARDRWGWRBAGRGGARRRBRKRKRDADDDDDAT 782
Oy	124 A---TATGGCGTANGTAACCTTCACAGAGAGGGCAGAGATGGAAGGGGGCTTGGAAC 180  :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : Db WTTWTTTTTRTDMDKMKTDTWTRMAARVWDRDDDDRRAAGTGRXMRRTWKMKRR 722
Oy	181 AAGAGAAATGTTCTCAAAATTCAGAAACTAGATGATTCACAGATGGGTCACTTCCT 240  :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : Db DTRWMDAADADPTANDRRRRRDDGDADAKKKGKRRRRBRATWDRTDAMWADAAMWT 662
Oy	241 GTCCCTGAGGTGGTGTAAGCATGTAAGTCTTATGGAAGAAAAGATGCATGTTAGGA 300  :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : Db TDITDIDMDKDRRKRGARRRRRTTARAAMWMTWKAMDMAKMDKTADRDWDAADTW 602
Oy	301 TGAAGTGAAGGCTAAGCAACAAGGCCAAGGCCCTATATCATCTCTAAAATGCTTTT 360  :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : Db TDARAADDMWAKARWRARRRARARADRRTTGGKTTTATWTTWAARPAAMAWMAWT 542
Oy	361 TTGATGTCTCTTAATTCACAAATGCTTCCAAGATGTAGCACAGAAAAAAGAAC 420  :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: : Db TTATWTTTWTWTTTWTWTTTWTWTTTAAWMAAWMTATWMAWMAWMAWMAWMAAT 482
Oy	421 ATWAGGACTTACTGSGGTGCTTTATCTTAAGCCTTGATCTGCTTTTACACAGCTTA 480

[illegible]

RESULT	21
LOCUS	AQ021826/c
DEFINITION	AQ021826 377 bp DNA linear GSS 18-SEP-1996 HS_3118_B2_G08_MR CTT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3118 Col=16 Row=N, genomic survey sequence.
ACCESSION	AQ021826
VERSION	AQ021826.1
KEYWORDS	GI:3624027
SOURCE	GSS.
ORGANISM	Homo sapiens (human)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 377) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED	10449764
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence tagged Connector  
Plate: 3118 row: N column: 16  
Class: BAC ends  
High quality sequence stop: 377.  
Location/Qualifiers  
1..377

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="P1ate=3118 Col=16 Row=N"
/sex="male"
/clone_11b="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelOAc11; BAC Clones in
E-Coli DH10B"

```

	Query Match	2.4%	Score 48.0	DB 9	Length 377
	Best Local Similarity	63.2%	Pred. No. 0.38		
Matches	74	Conservative	0	Mismatches 43	Indels 0
					Gaps 0
QY	55	TATAGTTGAGGACATGATGGATGCAAGTGCACGAAGTCAATTCATTAACCAAGTAAGTCA	114		
DB	238	TATTCATCATGCTCTGCATGCTGTAATATAGAAAACNCATATTCAGCTGGCTTAAGTATA	179		

**OY**      |||||  
115 AAGGAAATATATTGGTCATGTAAACCTTCA CAGAGAGGGCAGAGATNGAAGGGGC 171  
**Db**      |||||  
178 AAGGAAATTGATGAGGCCCATCAATTCAGCCACAGGTAGGGCAGAAGAGAGACTGGC 122

RESULT	22
AJ925404	
LOCUS	1062 bp mRNA linear EST 10-JUN-2005
DEFINITION	AJ925404 Theileria annulata merozoite Theileria annulata cDNA clone ta005f04_gtk, mRNA sequence.
ACCESSION	AJ925404
VERSION	AJ925404.1 GI:67495771
KEYWORDS	EST.
SOURCE	Theileria annulata
ORGANISM	Theileria annulata

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 10,62)	Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C.A., Weir, W., Kehonou, A., Aslett, M., Bishop, R., Boucher, C., Cochet, M., Coulson, R.M.R., Cronin, A., de Villiers, E., Fraser, A., Foster, N., Gardner, M., Goble, A., Griffiths-Jones, S., Harris, D.E., Kater, F., Lake, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F., Nene, V., O'Neill, S., Price, C., Quail, M.A., Rabinowitsch, E., Rawlings, N.D., Rutter, S., Saunders, D., Seeger, K., Shah, T., Squares, R., Squares, S., Tivey, A., Walker, R., Woodard, J., Dobbelaere, D.A.E., Langsley, G., Rajandream, M.-A., McKeever, D., Shiels, B., Tait, A., Barrett, B. and Hall, N.	The genome of the host-cell transforming parasite <i>Theileria annulata</i> and a comparison with <i>T. parva</i>		
	Unpublished (2005)			Contact: Pain A

```

Genome Campus, CH10 1SA, UNITED KINGDOM
Merozoite cDNA library: Frank Katzev and Brian Shields, Division of
Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
Location/Qualifiers
1..1062
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"
/clone="tam005f04_g1k"
/dev_stage="merozoite"
/lab_host="Bos taurus (Cow)"
/clone_lib="Theileria annulata merozoite"
/note="country: Turkey;Ankara"

```

	Query Match	Best Local Similarity	2.4#; Score 48.8; DB 1; Length 1062; Pred. No. 0.51; Matches 133; Conservative 0; Mismatches 140; Indels 1; Gaps 1;
QY	1493	TATTGACCAATGGAATACAGAGCTCTTTCTCCACAGGCTCATATGTAAGAATTCATT	1552
DB	765	TATAAATAAANNNTATATTTATNNNTTTTAAACAGAAATTTATTTTATATCTTTTTT	824
QY	1553	AGATGGCTGAATAGACTGATCTGTCATCTCTGCTGCTCATTCATCAATAGAGAGTCAT	1612
DB	825	TATTTTAAATTAATATATTTTCATATTTTATTTTTTTGTTAAATTTNNCNTTAACCTT	884
QY	1613	TAGCTAAGAACAAAACTACAAATCTATGTAAATTAGAAGACAA-GCTGGTTTGTCTCA	1671
DB	885	TNCTTATGACTTTAAATTTAATTTTACNTTNTATACAGATATATCTGTTTACACTCAT	944
QY	1672	TATTAATAATAGAAAAAGAAACATCGTGAAGTCAAAATTTGTTTAACTAGGCTCATGG	1731
DB	945	ATTACTCTTTAAATTAATATATATAATATGTTAAATTATAGTNTTATTTTCACTTGGCAGGCT	1004
QY	1732	AGAACTATTAAAAAGATATTGGAATCTTTATGCA	1765
DB	1005	ACAATATTTTATNTGTATTTTATTTTATATGA	1038

FEATURES	source
LOCUS	CL105409/c
DEFINITION	CL105409.1 ISB1 Xenopus tropicalis genomic survey sequence.
ACCESSION	CL105409.1
VERSION	GI:40599044
KEYWORDS	GSS.
SOURCE	Xenopus tropicalis (western clawed frog)
ORGANISM	Xenopus tropicalis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus; Silurana.
AUTHORS	1 (bases 1 to 1254)
TITLE	Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
JOURNAL	Mardis, E. and Wilson, R.
COMMENT	A physical map of the xenopus tropicalis genome Unpublished (2003) Contact: Richard K Wilson Genome Sequencing Center Washington University School of Medicine Email: submissions@wustl.edu Insert Length: 75000 Std Error: 0.00 Seq primer: Sp6 ATTATGGTGACACTATAG Class: BAC ends High quality sequence start: 54 High quality sequence stop: 443. Location/Qualifiers 1..1254 /organism="Xenopus tropicalis" /mol_type="genomic DNA" /db_xref="taxon:8364" /clone="ISB1-44M8" /clone_id="ISB1" /note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN	
Query Match	2.4%; Score 48.8; DB 10; Length 1254;
Best Local Similarity	60.6%; Pred. No. 0.54;
Matches	80; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY	1624 CAAAACTACAACTCTATGTAATTAGAAAGACAAGCTGTTTCTCAATATATAAATATAG 16833
DB	998 CAGTTACAAAATAATATACGAATATAGAAAAAATTTGTAATTTGAAATATATATAAAGAA 939
QY	1684 AAAAAAGAACATCTGGAAGTCAAAATATTTGTTTATATCAGTCATTGAGATCTATTTAA 17433
DB	938 GATATAAAGAAATTTTAAATATGTAATATTTTAAATATTAATAAAGATATTAATTTATTTGAC 879
QY	1744 AAAAGTATTTGAA 1755
DB	878 AAGTATTTAGAA 867
RESULT 24	
AQ338558/c	391 bp DNA linear GSS 12-JAN-1999
LOCUS	HS 3118 B1 G08 MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION	sapiens genomic clone Plate=3118 Col=15 Row=N, genomic survey
ACCESSION	AQ338558
VERSION	AQ338558.1
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 391)
	Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D. and Hood, L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Tel: (206) 616-3887  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 3118 row: N column: 15  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 391.  
Location/Qualifiers  
1. .391  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="plate=3118 Col=15 Row=N"  
/sex="male"  
/clone\_1db="CIT Approved Human Genomic Sperm Library D"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

Query Match 2.4%; Score 48.2; DB 9; Length 391;  
Best Local Similarity 63.2%; Pred. No. 0.55;  
Matches 74; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 55 TATAGTCAGGACCATGGATTCGACGAGACGAGAAATCTCAATTCAAACCAAGTAATGCTAA 114  
|||||  
DB 238 TATTCATCATGACTCTGACTGCTGTCATAGAAAAACCAATTCAGAGCTTGCTTAATATA 179  
|||||  
QY 115 AAGGAAATATATATGCTCATGTAACTTCATCAGACAGAGGCGAGATGCAAGGGGC 171  
|||||  
DB 178 AAGGAGAAATTGATGAGCCCATCAATCCATCCAGCGACAGGTAAGGCAAGAGAGAGTGGC 122  
|||||

RESULT 25  
CONS016E2/C  
LOCUS  
DEFINITION 1204 bp DNA linear GSS 26-JUL-1999  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CONS016E2 1204 bp DNA linear GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL106628  
AL106628.1 GI:5622852  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1204)  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CRPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.  
Location/Qualifiers

Query Match	2.4*	Score 48;	DB 10;	Length 958;
Best Local Similarity	50.4*	Pred. No. 0.8;		
Matches 117;	Conservative	0;	Mismatches 115;	Indels 0;
Gaps	0;			
Db	1540	TAAGAATTCATTGATTTGGCTGGAAAGAAGACGTATGTCATTTCTCTGCTCACTATC	1599	
Db	517	TAAGAAATTTAAAGAAAGATTAAGAAATTAATGATTAATTTGATTAATGATTAATTA	576	
Db	1600	ATAAGAAAGTCATTAGCTTAAGAAACAAAACTCAATCTATGTAATTAAGAAACAACT	1659	
Db	577	AAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	636	
Db	1660	CGTTTTCCTCAATTAATTAATTAAGAAAGAAACCATGCAAACTCAAAATTTTGT	1719	
Db	637	AAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	696	
Db	1720	ATCAGTCATTGGAATCTATTAAAGTATTTGAATCTTTATGATGAGA	1771	
Db	697	AAAAAGTAAATTAATTAATTTTAATTAAGAAATTAATTAATTAATTAATTA	748	
RESULT 27				
CNS020K7		1092 bp	DNA	linear
DEFINITION				GSS 01-SEP-2000
LOCUS				
DEFINITION				Tetradodon nigroviridis genome survey sequence T7 end of clone
DEFINITION				22L11 of library G from Tetradodon nigroviridis, genomic survey
DEFINITION				sequence.
DEFINITION				ALI75696
DEFINITION				ALI75696.1 GI:7813753
DEFINITION				GSS: genome survey sequence.
DEFINITION				Tetradodon nigroviridis
DEFINITION				Tetradodon nigroviridis
DEFINITION				Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DEFINITION				Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
DEFINITION				Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
DEFINITION				Tetraodontidae; Tetraodon.
DEFINITION				1
DEFINITION				Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
DEFINITION				Bernot,A., Fizames,C., Mincker,P., Broctier,P., Quehier,F.,
DEFINITION				Saurin,W., and Weissenbach,J.
DEFINITION				Estimate of human gene number provided by genome-wide analysis
DEFINITION				using Tetradodon nigroviridis DNA sequence
DEFINITION				Nat. Genet. 25 (2), 235-238 (2000)
DEFINITION				10835645
DEFINITION				2
DEFINITION				Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Coataz,C.,
DEFINITION				Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quehier,F.,
DEFINITION				Saurin,W., Bernot,A., and Weissenbach,J.
DEFINITION				Characterization and repeat analysis of the compact genome of the
DEFINITION				freshwater pufferfish Tetradodon nigroviridis
DEFINITION				Genome Res. 10 (7), 939-949 (2000)
DEFINITION				10899143
DEFINITION				3
DEFINITION				(bases 1 to 1092)
DEFINITION				Genoscope.
DEFINITION				Direct Submission
DEFINITION				Submitted (12-APR-2000) Genoscope - Centre National de Sequenage :
DEFINITION				BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
DEFINITION				- Web : www.genoscope.cns.fr)
DEFINITION				This sequence is a single read and was generated as part of a large
DEFINITION				scale clone-and-sequencing project of the Tetradodon nigroviridis
DEFINITION				genome. For more information, please take a look at
DEFINITION				http://www.genoscope.cns.fr/tetradodon.
DEFINITION				Location/Qualifiers
DEFINITION				1..1092
DEFINITION				/organism="Tetradodon nigroviridis"
DEFINITION				/mol_type="genomic DNA"
DEFINITION				/db_xref="taxon:99883"
DEFINITION				/clone="22L11"
DEFINITION				/clone_1b="G"
DEFINITION				/note="Genoscope sequence ID : COAG222CF06LPI
DEFINITION				end : T7"



## ORIGIN

Query Match 2.4%; Score 48; DB 10; Length 1092;

Best Local Similarity 37.9%; Pred. No. 0.83;

Matches 72; Conservative 39; Mismatches 79; Indels 0; Gaps 0;

QY 1596 TATCATAGAGACTCTTAGCTTAAGAACAACTACATCTATGTAATTAGAGACCA 1655

DB 627 WWWAAAAAAMTTTTTTTTTAAAMAAAAAMAAAAAMAAAAAMAAAAAMAAAA 686

QY 1656 AGCTGGTTTTCCTCATATAAATAAGAAAACCATGGAAGTCATAATTTTG 1715

DB 687 TTTTTTTTTTTTTTTTAAAAAAMAAAAAMAAAAAMAAAAAMAAAAAMAAAA 746

QY 1716 TTTAATCAGGTCATGAGATCTATTAAAGATTGATCTTATGATGAGACTAT 1775

DB 747 WWWAAWTMTTTTTTTTAAAMAAAAAMAAAAAMAAAAAMAAAAAMAAAAAMAAAA 806

QY 1776 CTGACTCAA 1785

DB 807 TTTAAAMAAA 816

RESULT 28

AL551397 950 bp mRNA linear EST 30-MAR-2004

LOCUS AL551397 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

ACCESSION AL551397

VERSION AL551397.3 GI:45856202

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Li,W.B., Gruber,C., Jesse,J., and Polayes,D.

AUTHORS Full-length cDNA libraries and normalization

TITLE Unpublished (2001)

JOURNAL On Feb 15, 2001 this sequence version replaced gi:31273213.

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 Evry cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

4987.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?cs=CS0DI064BD02QP1&amp;c=4987.r.

FEATURES

source

1. 950

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DI064YC04"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 2.4%; Score 47.8; DB 1; Length 950;

Best Local Similarity 36.5%; Pred. No. 0.9;

Matches 58; Conservative 42; Mismatches 59; Indels 0; Gaps 0;

QY 1597 ATATTAAGAAAGTCATTAGCTAAGAACTACATCTATGTAATTAGAGACCA 1656

DB 1718 TATCAGGTCATTGAGATCTATTAAAGTAATTGAAATCTTTATGATGAGA 1771

## ORIGIN

Query Match 2.4%; Score 47.6; DB 10; Length 1140;

Best Local Similarity 47.6%; Pred. No. 1.1; Indels 0; Gaps 0;

Matches 140; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1478 TACATATAAAGGGGATTTAGCAATGTGATACAGAGCTTTTCACAGATCTCATA 1537

DB 468 TATATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 527

QY 1538 TGTAAAGATTCATTAATGATGGCTGAATAGACTGATCTGTCATTTCTCTGCTCACTTA 1597

DB 528 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 587

QY 1598 TCATTAAGAGTCATTAGCTAAGAACCAAACTCAATCTATGTAATTAGAGAACAG 1657

DB 588 AATTAAATATTTAATATCTTTATTAACATATTTTAAATTAATTAATTAATTAATTA 647

QY 1658 CTGGTTTCTCAATATAAATAAATAAAGAAACCATGTGAAAGTCAAAATATTTGTT 1717

DB 648 AATTAATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 707

QY 1718 TATCAGGTCATTGAGATCTATTAAAGTAATTGAAATCTTTATGATGAGA 1771

RESULT 29

CL071552 1140 bp DNA linear GSS 31-DEC-2003

LOCUS CL071552 CH216-121C5 RM1.1 CH216 Xenopus tropicalis genomic clone

ACCESSION CH216-121C5, genomic survey sequence.

VERSION CL071552

KEYWORDS GSS.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM Xenopus tropicalis

REFERENCE Xenopus tropicalis

AUTHORS Xenopus tropicalis

TITLE Xenopus tropicalis

JOURNAL Xenopus tropicalis

COMMENT Xenopus tropicalis

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@wustl.edu

Insert length: 175000 Std Error: 0.00

Seq primer: RM1 TACGACTCATATAGCGAGA

Class: BAC ends

High quality sequence start: 43

High quality sequence stop: 164.

FEATURES

source

1. 1140

/organism="Xenopus tropicalis"

/mol\_type="genomic DNA"

/strain="Nigerian frog"

/db\_xref="taxon:8364"

/clone="CH216-121C5"

/sex="male"

/cell\_line="Stock 248 F7A2, inbred N7"

/clone\_lib="CH216"

/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis

BAC library"

ORIGIN

Query Match 2.4%; Score 47.6; DB 10; Length 1140;

Best Local Similarity 47.6%; Pred. No. 1.1; Indels 0; Gaps 0;

Matches 140; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1478 TACATATAAAGGGGATTTAGCAATGTGATACAGAGCTTTTCACAGATCTCATA 1537

DB 468 TATATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 527

QY 1538 TGTAAAGATTCATTAATGATGGCTGAATAGACTGATCTGTCATTTCTCTGCTCACTTA 1597

DB 528 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 587

QY 1598 TCATTAAGAGTCATTAGCTAAGAACCAAACTCAATCTATGTAATTAGAGAACAG 1657

DB 588 AATTAAATATTTAATATCTTTATTAACATATTTTAAATTAATTAATTAATTAATTA 647

QY 1658 CTGGTTTCTCAATATAAATAAATAAAGAAACCATGTGAAAGTCAAAATATTTGTT 1717

DB 648 AATTAATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 707

QY 1718 TATCAGGTCATTGAGATCTATTAAAGTAATTGAAATCTTTATGATGAGA 1771

[illegible]

FEATURES	Source	Location/Qualifiers
ORIGIN		1..1201 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACN15K02" /clone_lib="DrosBAC" /plasmid="pBelobAC11" /note="end : Sp6"
COMMENT		Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sege@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CERP (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
Query Match	2.4%	Score 47.4; DB 10; Length 1201;
Best Local Similarity	31.7%	Pred. No. 1.2;
Matches 137; Conservative	89;	Mismatches 203; Indels 3; Gaps 1;
QY	1362	TAAAGCTACACACACACACACCCCTCATGCAACACTGGCTTGGTAATACATATCTAT 1421
DB	1200	TMTMTATCTMTCTTCTGCHWTTTCTCTGCATCMTCTTTHYHTTCTCACAVMHTTTTY 1141
QY	1422	TATGAGAGCTGTGCTCTTAAGCGTAATATGTTTATATGACATPAGGCTCTTGCTTACA 1481
DB	1140	TACHAATTCYVMWTTTATMTATATAVAATHTTGMHAAMAWAACATATMTWMTA 1081
QY	1482	TATATAA---GGGATATTGAGCAATGTGATPACAGAACTTTTCTCCACAGTCTCATAT 1538
DB	1080	KATCAAAHHWMMMMMMMTTAAYYCYVMVYTTTAMTWTWTTTTATMMCYAAMWMCAMVTC 1021
QY	1539	GTAAGAAATTCATTAGATGTGGCTGAAPDAGACTGATGTCGTCATTCCTGCTCATTTAT 1598
DB	1020	YAMTMAAAYATMMHMMHMAATATATMTTCMTCTTTHHAACTCMTATMAACCMACAW 961
QY	1599	CATAGAAGACTATTAGCTAAGAGACAAACAACTACATCTATGTATATTAGAGAACAGC 1658
DB	960	AAATAAABAAWMACTHMMMMMMMMMAAAMMAAAMMAAAMMAAAMMAAAMMAAAMMAA 901
QY	1659	TGTTTGGTCTCATATATAAATAAAGAAAAGAAACCATGTGAAGTCAAAATATTTGTTT 1718
DB	900	AAAAAAAAMCCAAATAAACTAACCACTAAACAAAMMAAAMMAAAMMAAAMMAAAMMAA 841
QY	1719	AATAGAGCTATGGAATCTATTAAAGATTTGATTCCTTATATGATGAGAACTATCTT 1778
DB	840	YAAATATMTWTTTCACAPAAAMATGNBSMTTATATCRACCTCATYATMTTATATCACABT 781
QY	1779	GACTCAAGTGGA 1790
DB	780	CABTTACBKDA 769
RESULT 32		
CNS0161D		
LOCUS	1225 bp	DNA linear GSS 26-JUL-1999
DEFINITION		Drosophila melanogaster genome survey sequence sp6 end of BAC
		BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
		fly), genomic survey sequence.
ACCESSION		AL106171
VERSION		ALI06171.1 GI:5620504
KEYWORDS		GSS.
SOURCE		Drosophila melanogaster (fruit fly)
ORGANISM		Drosophila melanogaster.
		Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
		Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT	FEATURES	ORIGIN
1	(bases 1 to 1225)	Ephydroidea: Drosophilidae: Drosophila.				
2	Genoscope.					
3	Direct Submission					
4	Submitted (23-JUL-1999)	Genoscope - Centre National de Sequencage :				
5	BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr					
6	- Web : www.genoscope.cns.fr)					
7	Determination of this BAC-end sequence was carried out as part of a					
8	collaboration with the European Drosophila Genome Project (EDGP) -					
9	http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC					
10	library (Dros BAC) was made by Alain Billaud at CEPH (Centre					
11	d'Etude du Polymorphisme Humain) with funding provided by a MRC					
12	project grant. The DNA was prepared from embryos by Alain Bucheton					
13	and Genevieve Payan. It has been constructed in the vector					
14	pbeloBAC11.					
15	Location/Qualifiers					
16	1. .1225					
17	/organism="Drosophila melanogaster"					
18	/mol_type="genomic DNA"					
19	/db_xref="taxon:7227"					
20	/clone="BACN15C18"					
21	/clone_11b="DrosBAC"					
22	/plasmid="pbeloBAC11"					
23	/note="end : Sp6"					
24	Query Match	2.4%	Score 47.2;	DB 10;	Length 1225;	
25	Best Local Similarity	27.7%	Pred. No. 1.4;			
26	Matches 130;	Conservative 101;	Mismatches 239;	Indels	0;	Gaps 0;
27	OY	1286	TCTTGCTTTTTCACAGAGCCCTTATTTTACGTATATTTTGAATTCACGTATAT	1345		
28	DB	551	KNNKKKKKCTCKNNKCACTTTTWTATTTTWTATTTTWTATTTTWTATTTTWTATTAAT	610		
29	OY	1346	TTCTACATTTTTCCATTAAGTCATCTACACAAATACCTCTCAATGACAACTGGCTTT	1405		
30	DB	611	AAATTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	670		
31	OY	1406	GCTAATACATATCTATTATAGAGAGCTGCTTTTAAAGCGTAAATGTTTATATGACCTA	1465		
32	DB	671	TMTMHYTTTCCTCCYCCCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	730		
33	OY	1466	AGGCTCTGGCTCATATTTAAAGGGATTTGAGCAATGTGATACAGAGCTTTCTCC	1525		
34	DB	731	CCKYYGCMMAAAMAAAYAKCCBKKNKKKKMKKHAADAAAMAMNMAMNMAMNMAMNM	790		
35	OY	1526	ACAGGCTCATATGTAAGAATTCATTTGATTTGCTGAATAGACTGATCTGTCATTT	1585		
36	DB	791	WAAHAYNAAAAATTTAAATTCACAAAMMAAAAAAAMMAAAAAAAMMAAAAAAAMMA	850		
37	OY	1586	TCTGCTCATATCATAGAAGAGTCATTAGCTAAGAACAAAACTACAAATCTATGTAT	1645		
38	DB	851	KKKKDDKDAARARADKRAAAAAAATATATAAAAAAAMMAAAAAAAMMAAAAAAAMMA	910		
39	OY	1646	TAGAAGAACAAAGCTGTTTGTCTCATATTAATAAATAAGAAAAAAGAACCAATGTAAGTC	1705		
40	DB	911	AMAAAYAAATATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	970		
41	OY	1706	AAATATTTTGTATTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC	1755		
42	DB	971	WTATATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT	1020		
43	LOCUS	CL022041	1281 bp	DNA	linear	GSS 31-DEC-2003
44	DEFINITION	CH216-13E20_Sp6.1	CH216	Xenopus tropicalis	genomic clone	
45	ACCESSION	CL022041				
46	VERSION	CL022041.1	GI:40463854			
47	KEYWORDS	GSS				
48	SOURCE	Xenopus tropicalis	(western clawed frog)			
49	ORGANISM	Xenopus tropicalis				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.			
	1 (bases 1 to 1281)			
	Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.			
	A physical map of the xenopus tropicalis genome			
	Unpublished (2003)			
	Contact: Richard K Wilson			
	Genome Sequencing Center			
	Washington University School of Medicine			
	Email: submissions@watson.wustl.edu			
	Insert Length: 175000 Std Error: 0.00			
	Seq primer: Sp6 ATTGAGTGCACCTAATG			
	Class: BAC ends			
	High quality sequence start: 142			
	High quality sequence stop: 411.			
	Location/Qualifiers			
	1. 1281			
	/organism="Xenopus tropicalis"			
	/mol_type="genomic DNA"			
	/strain="Nigerian frog"			
	/db_xref="taxon:8364"			
	/clone="CH216-13E20"			
	/sex="male"			
	/cell_line="Stock 248 F7A2, indred N7"			
	/clone_id="CH216"			
	/note="Vector: pTRABAC2.1; CHORI-216 Xenopus tropicalis BAC library"			
ORIGIN				
Query Match	2.3%;	Score 46.8;	DB 10;	Length 1281;
Best Local Similarity	52.0%;	Pred. No. 1.8;		
Matches 129;	Conservative 0;	Mismatches 117;	Indels 2;	Gaps 1;
OY	1501	AATGATGATACGAGAGTCTTTCTCCACAGGCTCATATGTAAAGATTTCATTAGATTGGC	1560	
DB	497	AATTATTATTTAAAAAGGATTTTACAAATTAAGTATATATGTGCACATATTAATTAT--AT	554	
OY	1561	TGAATAGACTGATCTGTCCATTTCTCTGCTCACTTATCATAGAGGATCATTAGCTTAAG	1620	
DB	555	AGAAAGATATTAATTAATTCATTAATTAAGAAATTCATTTTTCATTACTAATCAACTATA	614	
OY	1621	GAACAAAACTACATCTATGTATTTAGAGACAGACTGGTTTGGTCAATATAAAAAT	1680	
DB	615	AATGAAAAACAATATTATTTATGTATTAATTTTCACATATTATTACATCATAGAAAAAGAT	674	
OY	1681	AAGAAAAAGAAACCATGTGAAAGTCAAAATTTTGTTAATCAGTCATTGAGAACTGAT	1740	
DB	675	AAAAATTAACCATTTATATATTAATTAATTAATTTTATTTATTAATAATTAAGCTAATTAATT	734	
OY	1741	TAAAGAT 1748		
DB	735	AAAGTAAAT 742		
RESULT 34				
LOCUS	DRO71573	511 bp	mRNA	linear EST 08-JUN-2005
DEFINITION	RTDK1_20_F06_g1_A029 Roosts, dark Pinus taeda cDNA clone			
ACCESSION	RTDK1_20_F06_A029 5', mRNA sequence.			
VERSION	DRO71573			
KEYWORDS	DRO71573.1 GI:670493321			
SOURCE	EST.			
ORGANISM	Pinus taeda (loblolly pine)			
REFERENCE	Pinus taeda			
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pratt, L., Cordonnier-Pratt, M.M., Lorenz, W.W., Zimmermann, C. and Dean, J.F.D.			
TITLE	An EST database from dark-treated loblolly pine (Pinus taeda) roosts			
JOURNAL	Unpublished (2005)			

## COMMENT

Other ESTs: RTDK1 20 F06.b1 A029  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 RNA prepared and library constructed by W. Walter Lorenz (School of Forest Resources, University of Georgia); plant material prepared by Craig Zimmermann (School of Forest Resources, University of Georgia) using rooted cuttings provided by the Forest Biology Research Cooperative (FBRC) and the CCLONES project at the University of Florida; sequencing done in the laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement  
 Seq primer: JENREV (CAGAAACAGCTATGACC).

## FEATURES

source

Location/Qualifiers  
 1..511  
 /organism="Pinus taeda"  
 /mol\_type="mRNA"  
 /strain="3 CCLONES"  
 /db\_xref="taxon:3352"  
 /clone="RTDK1 20 F06 A029"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Roots, dark"  
 /note="Organ: Root; Vector: pSL1180; Site 1: EcoRI; Site 2: XhoI; The library was prepared from poly(A) RNA from the roots of 1-year-old loblolly pine (Pinus taeda) cuttings that were rooted and then planted in washed sand. The rooted cuttings were maintained for 50 days (May 1 2003 harvest) under ambient conditions in a local greenhouse. They were kept on a weekly regimen of 0.5x nutrient-complete Hoagland's solution and supplemented with additional water sufficient to maintain a 15% soil moisture content. Twenty-four hours (24h) prior to harvesting roots for mRNA preparation, the potted trees were placed in a dark growth chamber at 25 C. Double-stranded cDNA was cloned unidirectionally into pSL1180. Inserts can be excised with EcoRI (5' end) and XhoI (3' end)."

## ORIGIN

Query Match 2.3%; Score 46.6; DB 8; Length 511;  
 Best Local Similarity 57.5%; Pred. No. 1.5;  
 Matches 103; Conservative 0; Mismatches 74; Indels 2; Gaps 1;

QY 1596 TATCATAGAGAGTCTAGCTAAGCAAAAAGTCAATCTATGTAATTAAGAGACA 1655  
 |||  
 DB 2 TACCAAGTAGAATTTTCATGATTAGGACCAATTCATACATCTTGTAAATTTAAACAA 61  
 QY 1656 AGTGGTTTGCATATATAAATAAGAAAGAAAGCAATGGAAGTCAAAATATATTG 1715  
 |||  
 DB 62 TACCATTTAATTAATAATTAATTAACCAAAAAACATTCACATGAAAAACAATTA 121  
 QY 1716 T--TTAATCAGCTCATTTGAGATCTATTAATAAGATTTGAATCTTATGATGAGAAC 1772  
 |||  
 DB 122 TAATTAATCCACCAAGAGATTCAGTCCCAAGTATCATTTTCATGATTAATGTGAAAC 180

RESULT 35  
 CV569142/c 617 bp mRNA linear EST 22-Oct-2004  
 LOCUS od01h09.y1 Human keratocornus cornea, unamplified, (od/oe) Homo  
 DEFINITION sapiens cDNA clone od01h09 5', mRNA sequence.  
 ACCESSION CV569142  
 VERSION CV569142.1 GI:54468336  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
 1 (bases 1 to 617)  
 AUTHORS Rabinowitz, Y., Dong, L. and Wistow, G.  
 TITLE Expressed sequence tag analysis of human keratocornus cornea  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 plate: 01 row: h column: 09

## FEATURES

source

Location/Qualifiers  
 1..617  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="od01h09"  
 /tissue\_type="Cornea"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human keratocornus cornea, unamplified, (od/oe)"  
 /note="Organ: Eye; Vector: pCMVSPORT6; Approximately 40ug total RNA was extracted from 7 adult human keratocornus corneas. A directionally cloned cDNA library in the pSPORT1 vector (Invitrogen) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetechn.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pCAGTAGTCTAGTACGCGAGCGCGCC(7)15-3']. cDNA was cloned in Not I/Sal I sites. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

## ORIGIN

Query Match 2.3%; Score 46.4; DB 7; Length 617;  
 Best Local Similarity 49.7%; Pred. No. 1.8; DB 171; Indels 2; Gaps 2;  
 Matches 171; Conservative 0; Mismatches 171;

QY 1448 AATGTTTATATGACCTAAGGCTCTTGCTTACATATAAAGGGATTGAGCAATGTGA 1507  
 |||  
 DB 617 AATGCTTTGTTTCTAAACCTGTATGTTTAAACAAATATGAAAGATGA 558  
 QY 1508 TACAGAAGTCTTTTCTCCACAGCTTCATATGAAGAATTCATTGATTTGCTGAATA 1567  
 |||  
 DB 557 ACCACAACAGTGTACCAAAAAGTCACTTGTTCACGACATCA -CTGTCTTTCACAC 499  
 QY 1568 GACTGATCTGTCCATTTCTGTCTCACTATCATAGAAGACATTAAGCAACAA 1627  
 |||  
 DB 498 AATTCAACTTAATAAGATTTAAGATTTAACTTCACTTAATGATTCAGAACATTA 439  
 QY 1628 AACTCAATCTATGTAATTAAGAACAAGCTGTTTGTCTCAATTAATAAATAAGAAA 1687  
 |||  
 DB 438 AAAAATCTCAATGATACATTTAATAATATGCTGCTTCATATATATTTCTTAAAAA 379  
 QY 1688 AGAACCATGTGA -AGTCAAAATATTTGTTTATATCAAGCTCATTTGAATTAATAA 1746  
 |||  
 DB 378 TGAATAACAGTAAAGGACACACTTTTATTAATCAACAACCTGAATATATATGTA 319  
 QY 1747 GATTTGAATTTCTTATGATGAGAACTATCTTGACCTCAAGTGA 1790  
 |||  
 DB 318 CTTTGTGAATTTTCTTTCTTTAGACATTTTCTCTAGATGA 275

RESULT 36  
 BH712636 644 bp DNA linear GSS 20-FEB-2002  
 LOCUS BH712636

DEFINITION BOMFC43TF BO\_2.3 KB Brassica oleracea genomic clone BOMFC43,  
genomic survey sequence.  
ACCESSION BH712636  
VERSION BH712636.1 GI:18804083  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea  
REFERENCE 1 (bases 1 to 644)  
AUTHORS Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,  
Uteirback,T.R., Wortman,J.R., White,O.R. and Town,C.D.  
TITLE Whole genome shotgun sequencing of Brassica oleracea and its  
application to gene discovery and annotation in Arabidopsis  
JOURNAL Genome Res. 15 (4), 487-495 (2005)  
PUBMED 15805490  
COMMENT Other\_GSSs: BOMFC43TR  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
1.644  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOMFC43"  
/clone\_1lb="BO\_2.3 KB"  
/note="Vector: pHOST1; Site 1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHOST1 using BstXI linkers"

ORIGIN  
Query Match 2.3%; Score 46.4; DB 9; Length 644;  
Best Local Similarity 53.3%; Pred. No. 1.9; Mismatches 86; Indels 0; Gaps 0;  
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 1597 ATCATAGAGAAAGCTTAGCTTAAGAACAAATCTAATCTATGTATATAGAAACAA 1656  
DB 451 ACCGAAGGAAATATCAAGAGATGAAAAATTCATAAATTCATTAATAAATAAGAG 510  
QY 1657 GCTGCTTTTGCCTCAATATATAAATAAGAAAAAGAACCATGTGAAGTCAAAATTTTGT 1716  
DB 511 AAACATTTTATTTAATTCATAGAAAGCAAAACATTAATATGTATTTTGAAGATGTC 570  
QY 1717 TTAATCAGCTCATAGAAATCTATTAAGAATTTGAATCTTTATGATGAGAACTATC 1776  
DB 571 AAATTTATCTCATATATGTTCAATCAAACTCTTTTAAATGTTGATGATGTGATC 630  
QY 1777 TTGA 1780  
DB 631 CCGA 634

RESULT 37  
AZ899171 649 bp DNA linear GSS 05-MAR-2001  
LOCUS RPCI-24-213116.TJ RPCI-24 Mus musculus genomic clone  
DEFINITION RPCI-24-213116, genomic survey sequence.  
ACCESSION AZ899171  
VERSION AZ899171.1 GI:13218116  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 649)  
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintrel,B., Levins,M.,  
Tsegaye,G., Geer,K., Krol,M., Shvartbeyn,A., Gebregeorgis,E.,  
Russell,D., de Jong,P. and Fraser,C.M.  
TITLE Mouse BAC End Sequences from Library RPCI-24  
JOURNAL Unpublished (1999)  
COMMENT Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderinfoframe.html>). BAC end  
page: [http://www.tigr.org/cdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 213 row: 1 column: 16  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
1.649  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-213116"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/clone\_1lb="RPCI-24"  
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;  
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
library was cloned in the pTARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

ORIGIN  
Query Match 2.3%; Score 46.4; DB 9; Length 649;  
Best Local Similarity 52.5%; Pred. No. 1.9; Mismatches 111; Indels 1; Gaps 1;  
Matches 124; Conservative 0; Mismatches 111; Indels 1; Gaps 1;  
QY 1595 TTATCATAGAGAGCTTAGCTTAAGAACAAATCTAATCTATGTAT- TAGAAGAA 1653  
DB 326 TAAACGAGTAGAAGCACTTAAGAGAAACACAAACATCAACAAACAGGTGAAGAAAA 385  
QY 1654 CAAGCTGCTTTGCCTCAATATATAAATAAGAAAAAGAACCATGTGAAGTCAAAATTTT 1713  
DB 386 GAACAAACCATTCATATCTTAATAATGAAAAAAGAAACATGAAGAATCACAAAAGA 445  
QY 1714 TGTTTATCAGGCTCATAGATCTATTAAGAATTTGAATCTTTATGATGAGAACT 1773  
DB 446 AGATTAACCTGTGTGGAAAACTTAGAAGAGATCAGAGTCATAGATCGACATCA 505  
QY 1774 ATCTTACTCAAGTGACAGATGAGTGGCTTTTGGCTGTGCTCCCTACGTAGAAA 1829  
DB 506 CCAAAAAATATACAGATACAGAGAAGAGATCTTAGGTGCAAGAAGATACCATAGAAA 561

RESULT 38  
AZ899889 733 bp DNA linear GSS 05-MAR-2001  
LOCUS RPCI-24-213J16.TJ RPCI-24 Mus musculus genomic clone  
DEFINITION RPCI-24-213J16, genomic survey sequence.  
ACCESSION AZ899889  
VERSION RPCI-24-213J16, genomic survey sequence.  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

**AUTHORS** Zhao, S., Niemann, W., Malek, J., Shatsman, S., Akhmer, B., Levits, M.,  
Tsegaye, G., Geer, K., Krol, J., Shvartsbeyn, A., Gebreyeorgis, E.,  
Russell, D., de Jong, P., and Frazer, C.M.  
**TITLE** Mouse BAC End Sequences from Library RPCT-24  
**JOURNAL** Unpublished (1999)  
**COMMENT** Other\_GSSs: RPCT-24-213016.TV

Department of Eukaryotic Genomics  
The Institute for Genomic Sciences  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPc1-24. For BAC library availability, please contact Pieter de Jong (pdjong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC endplate: [http://www.tigr.org/cdb/bac\\_ends/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/bac_end_intro.html).  
Plate: 213 row: J column: 16  
Seq primer: 596  
Class: BAC ends

FEATURES	Location/Qualifiers
source	1. .733

```

1. .733
/organism="Mus musculus"
/mol_type="Genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCT-24-213J16"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_1id="RPCT-24"
/clone_2id="RPCT-24"
/notes="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCT-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the pPARBAC1 cloning vector at the
BamHI sites using MhoI partially digested male C57BL/6J
DNA."

```

## ORIGIN

Query Match	2.3%	Score 46.4;	DB 9;	Length 733;
Best Local Similarity	52.5%;	Pred. No. 1.9;		
Matches 124;	Conservative 0;	Mismatches 111;	Indels 1;	Gaps 1;

OY	1595	TTATCATAGAGACATTAGCTAGAGAAACAAAATCTCAATCTGTATAT-TAAGAGA	1655
Db	365	TAAACGCTAGAGACCTTTAAAGAGAAACACAAACACATCAACAGGTGAAGAAAA	424
OY	1654	CAAGCTGGTTTGTCTCAATATATAAATAAGAAAAAACCATGTAAAGTCAAAATAT	1713
Db	425	GAAACAAACCATCCAAATATCTTAAATATGAAAAAAGAAACATGAAGAAATCACAAAAG	484
OY	1714	TGTTTAATCAGTCAATTGGAATCTATTAAGAAATTTGAATCTTTATATGTGAAC	1773
Db	485	AGATTAACCTGTGTGTGGAAAAACCTAAGAAAAAGATACGAGATCTAAGTGCAGCATCA	544
OY	1774	ATCTTGACTCAAGTGGACGTGTAGCTTTTGTGCTGTGTGCTCCATCGTAGAA	1829
Db	545	CCAAAAATACAGATACAGAAAGAGAAATCTTAGGTGCAGAGATACCATATGAAA	600

RESULT 39	LOCUS	DEFINITION
CNS0155H	1001 bp	DNA linear GSS 26-Jul-1999
CNS0155H		Drosophila melanogaster genome survey sequence SP6 end of BAC
		BACN1C23 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION VERSION	KEYWORDS	SOURCE	ORGANISM
AL105023	GI:5617037	GSS.	
AL105023.1			
		<i>Drosophila melanogaster</i> (fruit fly)	
		<i>Drosophila melanogaster</i>	
		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota	
		Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
		Ephyroidea; Drosophilidae; Drosophila.	

REFERENCE 1 (bases 1 to 1001)  
AUTHORS Genoscope,  
TITLE Direct Submission  
JOURNAL Submitted (23-UTL-1999) Genoscope - Centre National de Sequencage

**COMMENT**

Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES	Location/Qualifiers
source	1. .1001

```

organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN13C3"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : Sp6"

```

## ORIGIN

Query Match	2.3%;	Score 46.4;	DB 10;	Length 1001;
-------------	-------	-------------	--------	--------------

Best Local Similarity 38.7%; PRed.NO. 2.1;  
Matches 99; Conservative 46; Mismatches 106; Indels 5; Gaps 1;

b  
y  
y  
y  
b

1535 ATATGTGAAAAGAAATTCATTGATGGCTGAATAAGCATCATGTGCATTTCTCGCTAC 1594  
|||||:::|||||  
720 ATTAARWAAARAARABAAAAAATTTAWTAAWWMAWMAWMAAATAAAADDTTWATMBA 779  
|||||:::|||||

y 1595 TTATCATTAAGAAGTCATTGCTAAGAACAAAACTCAATCTATGTAAATTAGAGAAC 1658  
|||||:::|||||  
b 780 TAAAAAAAATWTAAAAAAATATTTAAWMAWMAAAATWAAAAATTTAAAAAAMWTAAWMAAAA 839  
|||||:::|||||

840 A A A T T A A T T T T T T A A A W A A A W A A T W A T A T A A A A D A A A A A A A A A A W A A W A T A A A A T 899

[illegible]

RESULT 40  
G757986

Accession	Length	Library	Species	Genome
CG575986	1364 bp	DNA	linear	GSS 24-OCT-2003
P053-2-DD2.zb	Ppa	ECORI	BAC Library	Pristionchus pacificus genomic,
DEFINITION	genomic survey sequence.			

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
1 (bases 1 to 1364)  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

## AUTHORS

**TITLE** An integrated physical and genetic map of the nematode *Pristionchus pacificus*

# ТОПКА

JOURNAL MOL. Genet. Genomics 269 (5), 715-722 (2003).

## COMMENT

Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany





Echinoidea; Euechinoidea; Echinacea; Echinoida;  
Strongylocentrotidae; Strongylocentrotus.  
1 (bases 1 to 889)  
Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and  
Ettensohn,C.A.  
A large scale analysis of mRNAs expressed by primary mesenchyme  
cells of the sea urchin embryo  
Development 128 (13), 2615-2627 (2001)  
JOURNAL  
PUBMED  
11493577  
COMMENT  
Contact: Erica Sodergren  
Human Genome Sequencing Center  
Baylor College of Medicine  
One Baylor Plaza, Houston, TX 77030, USA  
Tel: 713-798-7676  
Fax: 713-798-6977  
Email: ericas@bcm.tmc.edu  
NCBI Trace Archive: 496943148  
Insert Length: 1750 Std Error: 0.25  
Plate: 77 row: 0 column: 21.  
FEATURES  
source  
1..889  
/organism="Strongylocentrotus purpuratus"  
/mol\_type="mRNA"  
/db\_xref="taxon:7668"  
/clone="PMCSRP2-77021"  
/issue\_type="embryo"  
/cell\_type="primary mesenchyme cells"  
/lab\_host="E.coli"  
/clone\_idb="Sea Urchin primary mesenchyme cell cDNA  
library"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: MluI; oligo  
dt priming from poly A+ RNA, directionally cloned"  
ORIGIN  
Query Match 2.3%; Score 46; DB 8; Length 889;  
Best Local Similarity 50.0%; Pred. No. 2.6;  
Matches 115; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
QY 1482 TATTAAGGGGTTATGAGCATGTGATACAGAACTTTTCTCCACAGTCTCATATGTA 1541  
DB 246 TTTTAATGCCCAAAAAAATCCCAAAAACCAACCCCTCCCGGTTTATTTA 187  
QY 1542 AAGATTCATTAAGTGGCTGGAATATAGCATGATCTGCCATTTCTGCTCACTATCAT 1601  
DB 186 AAAAACTTTTTCATCAAAAAAATTTAAATTTTCCCGCCAGAAAAA 127  
QY 1602 AAGCAAGTCATTCCTAAGACCAAACTACATCATGTAAATTGAAGACAGCTGG 1661  
DB 126 AA 67  
QY 1662 TTTTGCTCATATTAATAATAGAAAAAGAAACCATGTGAAGTCAATA 1711  
DB 66 TTTAAAGAAAAAATTTTAAAAAAAAAAAAAAAAAAAAAAAAAATTAAGA 17  
RESULT 43  
CNS000DKY 928 bp DNA linear GSS 04-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
DEFINITION BACR27A24 of RPL-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL071865 GI:4948170  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephyroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 928)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutopo Osogawa and  
Aaron Mammosser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPL-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
FEATURES  
source  
1..928  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR27A24"  
/clone\_idb="RPL-98"  
/note="end : T7"  
ORIGIN  
Query Match 2.3%; Score 46; DB 10; Length 928;  
Best Local Similarity 30.2%; Pred. No. 2.6;  
Matches 110; Conservative 74; Mismatches 180; Indels 0; Gaps 0;  
QY 1409 AATACATATCTATATGAGAGCTGCTCTTAAAGCGTAATGTTTATATGACTAAGG 1468  
DB 838 MAMMMAMMMAMMMTTTTHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 779  
QY 1469 CTTCTGGCTTACATATAAAGGGTATGAGCAATGATGACGATCTTCTCCACA 1528  
DB 778 TTTTMMTMMTMMTMMTMMNNNNMCHMAMTMMTMMTMMTMMTMMTMMTMMTMMTMM 719  
QY 1529 GGTCTCATATGTAAGAATTCATTAAGTGGTGAATGAGTATCTGCTCATTTCTCT 1588  
DB 718 AAAAAAAAAAATTAATTTTWTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 659  
QY 1589 GCTCACTTATCATAGAGTCAATGCTAAGACAAAACCTACATCTATGTAATGAG 1648  
DB 658 TWTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 599  
QY 1649 AAGAACAGCTGTTTGTCTCAATTAATAAATGAAGAAACCATGTGAAGTCAAA 1708  
DB 598 AATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 539  
QY 1709 ATATTTGTTTAATCAGGTCATTAATTAATTAATTAATTAATTAATTAATTAAT 1768  
DB 538 WAAAAAATTTTWTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 479  
RESULT 44  
CNS01038 1201 bp DNA linear GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION BACN03017 of Drosophila library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL098462 GI:5610073  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;







```
Db      39238  TGAATAGACTGATCTGTCCATTTCTGCTCACTTATCAATAGAGTATTGCTAG 39297
Qy      1621  GAACAAAACTACATCTATGTATTTAGAGAAACAGCTGGTTTGTCTCAATATAAAAT 1680
Db      39298  GAACAAAACTACATCTATGTATTTAGAGAAACAGCTGGTTTGTCTCAATATAAAAT 39357
Qy      1681  AAGAAAAAGAACCATGTGAAGTCAAAATATTGTTTATCAGGTATTGAGAACTAT 1740
Db      39358  AAGAAAAAGAACCATGTGAAGTCAAAATATTGTTTATCAGGTATTGAGAACTAT 39417
Qy      1741  TAAAAAGTATTGAAATCTTTATGATGAGAACTATCTGAAGTGAAGTGAAGTGA 1800
Db      39418  TAAAAAGTATTGAAATCTTTATGATGAGAACTATCTGAAGTGAAGTGAAGTGA 39477
Qy      1801  CTTTTTGCCCTGTGTGCTCCCTACGTGAAGAGGCTTGTCAATAAGTCTTATATG 1860
Db      39478  CTTTTTGCCCTGTGTGCTCCCTACGTGAAGAGGCTTGTCAATAAGTCTTATATG 39537
Qy      1861  AGGTGCCAAGTTAAGTCCCAAGCTTCTTAAAGCATACAGATTTGTTTAACT 1920
Db      39538  AGGTGCCAAGTTAAGTCCCAAGCTTCTTAAAGCATACAGATTTGTTTAACT 39597
Qy      1921  TTTAGTGAAGTGAAGGAAATTAACAATCCCTCTGGAGAACTTCTCTCATCTTGT 1980
Db      39598  TTTAGTGAAGTGAAGGAAATTAACAATCCCTCTGGAGAACTTCTCTCATCTTGT 39657
Qy      1981  GAAAGTCAATTTCTCCAGAAATTC 2001
Db      39658  GAAAGTCAATTTCTCCAGAAATTC 39678

RESULT 2
LOCUS   AF222855 80622 bp DNA linear PRI 17-Jul-2000
DEFINITION Homo sapiens clone HC chromosome 10 map 10q25.2 genomic sequence.
ACCESSION AF222855
VERSION   AF222855.1 GI:9246845
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo
1 (bases 1 to 80622)
Barry,A.E., Bateman,M., Howman,E.V., Cancilla,M.R., Tainton,K.M.,
Irvine,D.V., Saffery,R. and Choo,K.H.
The 10q25 neocentromere and its inactive progenitor have identical
primary nucleotide sequence: further evidence for epigenetic
modification
JOURNAL Genome Res. 10 (6), 832-838 (2000)
PUBMED 10854414
REFERENCE 2 (bases 1 to 80622)
AUTHORS Barry,A.E.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-2000) Chromosome Research Unit, The Murdoch
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
Parkville, Melbourne, Victoria 3052, Australia
REMARK Genomic sequence from human 10q25.2, clone1b=HC
FEATURES
source
1.80622
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25.2"
/clone="HC"
/repeat_region complement(3..176)
/rpt_family="Alu"
/repeat_region complement(979..3564)
/rpt_family="Tigge1"
2967..3094
/rpt_family="Alu"

repeat_region 3664..3934
/rpt_family="Alu"
complement(4920..5181)
/rpt_family="Alu"
repeat_region complement(6287..6579)
/rpt_family="Alu"
repeat_region 6625..6966
/rpt_family="MLT2B2"
complement(9215..9490)
/rpt_family="Alu"
repeat_region complement(11280..11565)
/rpt_family="Alu"
repeat_region complement(12428..12675)
/rpt_family="Alu"
repeat_region complement(14304..14583)
/rpt_family="Alu"
repeat_region 17503..17994
/rpt_family="Alu"
complement(20250..20533)
/rpt_family="Alu"
repeat_region complement(23031..23323)
/rpt_family="Alu"
repeat_region complement(24301..24479)
/rpt_family="MIR"
24951..25250
/rpt_family="Alu"
complement(25987..26342)
/rpt_family="TBE1"
27097..27335
/rpt_family="MER2"
27538..27814
/rpt_family="Alu"
27891..28161
/rpt_family="Alu"
29063..29334
/rpt_family="Alu"
complement(31025..31093)
/rpt_family="MLT1"
31817..32352
/rpt_family="MER1"
32489..32607
/rpt_family="MLT1"
33039..33316
/rpt_family="Alu"
33469..33760
/rpt_family="Alu"
complement(34527..34590)
/rpt_family="MIR"
34689..34995
/rpt_family="Alu"
complement(36349..36437)
/rpt_family="MIR"
38796..39079
/rpt_family="Alu"
39441..39693
/rpt_family="Alu"
complement(43176..43237)
/rpt_family="MIR"
complement(45203..45478)
/rpt_family="Alu"
45892..46025
/rpt_family="MIR"
complement(46618..46851)
/rpt_family="Alu"
48623..48939
/rpt_family="Alu"
complement(49048..50698)
/rpt_family="L1"
complement(50710..50972)
/rpt_family="Alu"
complement(51031..51192)
/rpt_family="L1"
complement(51450..51739)
```

```
repeat_region /rpt family="Alu"
complement(53770. .53562)
/rpt family="Alu"
54261. .54808
repeat_region /rpt family="LTR12"
59460. .60054
repeat_region /rpt family="LTR12"
60349. .60633
/rpt family="Alu"
complement(60669. .60801)
/rpt family="MIR"
complement(61684. .61852)
/rpt family="MERS"
63466. .63753
repeat_region /rpt family="Alu"
64063. .64228
/rpt family="MERS"
64446. .64521
/rpt family="MIR"
65794. .65883
repeat_region /rpt family="MERS"
66772. .66951
/rpt family="MER20"
complement(67895. .68416)
/rpt family="MER41"
complement(73192. .73470)
/rpt family="Alu"
74086. .74367
repeat_region /rpt family="Alu"
75316. .76183
/rpt family="Alu"
complement(77771. .78059)
/rpt family="Alu"
```

## ORIGIN

Query Match	100.0%;	Score 2001;	DB 8;	Length 80622;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2001; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

OY	1	ATCTTATTTGTAGAGAAAAAGAAAAAGAGAGTGAATAATATATACGTGCATATAT	60
Db	78622	ATCTTATTTGTAGAGAAAAAGAAAAAGAGAGTGAATAATATATACGTGCATATAT	78681
OY	61	TCAGGACCATGGATTGGCAAGTGCAGAAACTCAATTCAACTCAACGTAAAGTCAAAAGAA	120
Db	78682	TCAGGACCATGGATTGGCAAGTGCAGAAACTCAATTCAACTCAACGTAAAGTCAAAAGAA	78741
OY	121	AATATATTGGCTCATGTAACTCTCTCAAGAGAGGACGAGTGAAGGAGGCTTTGGGAAC	180
Db	78742	AATATATTGGCTCATGTAACTCTCTCAAGAGAGGACGAGTGAAGGAGGCTTTGGGAAC	78801
OY	181	AAGAGAAATGTTCTCCAANTTCTAGGAATACTAGATTATAGTCCAGATGGGTCACTTCC	240
Db	78802	AAGAGAAATGTTCTCCAANTTCTAGGAATACTAGATTATAGTCCAGATGGGTCACTTCC	78861
OY	241	GTCCCTGAGGGTGGTGTAGCGATGGTAGACTTTATGGAGAAAGATGCATGTTAGGA	300
Db	78862	GTCCCTGAGGGTGGTGTAGCGATGGTAGACTTTATGGAGAAAGATGCATGTTAGGA	78921
OY	301	TGAAGTGAAGGCTTAAGCAAAACAAGGGCAAGGGCCATATATCATGCTAAAAATGTTTTT	360
Db	78922	TGAAGTGAAGGCTTAAGCAAAACAAGGGCAAGGGCCATATATCATGCTAAAAATGTTTTT	78981
OY	361	TTTATATGCTCTCTTAATTTCAAAATGCTTCCAAACAAATACACACAGGAAAAAGAAC	420
Db	78982	TTTATATGCTCTCTTAATTTCAAAATGCTTCCAAACAAATACACACAGGAAAAAGAAC	79041
OY	421	ATTAGGAACTACTAGTGGGTGCTTTATCTTAAAGCTTGACTTGCTTTCAACACTTA	480
Db	79042	ATTAGGAACTACTAGTGGGTGCTTTATCTTAAAGCTTGACTTGCTTTCAACACTTA	79101
OY	481	CTCATGCTTATACCTGAAGCATATATGCTCTGTAAGGCTTGTCAAGGCTTTCTACTAAG	540

Db	79102	CTCACTCTCTGTACCTGAGGCCCATAGCCCTGTAAAGAGTTCTGACAGGGTTTCTACTAAG	79161
Qy	541	CTGGGTTCTTTATATAGGCTCTCTCCATTTCTGTGCTCTCACTCTATGATCTTTCTCTT	600
Db	79162	CTGGGTTCTTTATATAGGCTCTCTCCATTTCTGTGCTCTCACTCTATGATCTTTCTCTT	79221
Qy	601	TTTCTCACTCTGGGACCTGGTGGCTGTTTGTATANGACCTGCTTAGCTTTGCTTGGGTTT	660
Db	79222	TTTCTCACTCTGGGACCTGGTGGCTGTTTGTATANGACCTGCTTAGCTTTGCTTGGGTTT	79281
Qy	661	TTTTCTGGGGACAATGCTCTTCAATTAATCCTAGACCAAAATTAACCTACAGCCACTGGGCCA	720
Db	79282	TTTTCTGGGGACAATGCTCTTCAATTAATCCTAGACCAAAATTAACCTACAGCCACTGGGCCA	79341
Qy	721	GGCTCTCTCTCTCCACACTGGACCAATGTTCCAGGGCTCTTCACTTAGTTAGTCAAG	780
Db	79342	GGCTCTCTCTCTCCACACTGGACCAATGTTCCAGGGCTCTTCACTTAGTTAGTCAAG	79401
Qy	781	CATTCTTGGGAAAAAGAGCCATAGTTAACATATAGCACTTCTAGCAATAGTAATCTTTTGG	840
Db	79402	CATTCTTGGGAAAAAGAGCCATAGTTAACATATAGCACTTCTAGCAATAGTAATCTTTTGG	79461
Qy	841	ACATGTTGTAAAGATCTATTACATTTTGTATATTAAGCACTTCCCTATGGAACAACACAC	900
Db	79462	ACATGTTGTAAAGATCTATTACATTTTGTATATTAAGCACTTCCCTATGGAACAACACAC	79521
Qy	901	GAACATAAGCTGCTCTGGAAATGACAGGGTGGCTCTCAATATAGATGTTCTTAGAGCT	960
Db	79522	GAACATAAGCTGCTCTGGAAATGACAGGGTGGCTCTCAATATAGATGTTCTTAGAGCT	79581
Qy	961	GTATTTTGGGCACTTAACTATTTCTCCACTTAAGGGCAAGACATGAAATTAACACACAC	1020
Db	79582	GTATTTTGGGCACTTAACTATTTCTCCACTTAAGGGCAAGACATGAAATTAACACACAC	79641
Qy	1021	TAAGTTTGTCAATGTCATGTAGTAGTCTCAGGCACTGACGCTCAGAGTGAACCTGAC	1080
Db	79642	TAAGTTTGTCAATGTCATGTAGTAGTCTCAGGCACTGACGCTCAGAGTGAACCTGAC	79701
Qy	1081	CTCTTATGTGTGTCAGACCCTTTCTTCCCTCAGAAGTCACTGTGTTTCTGTGACTCTC	1140
Db	79702	CTCTTATGTGTGTCAGACCCTTTCTTCCCTCAGAAGTCACTGTGTTTCTGTGACTCTC	79761
Qy	1141	CATAGGAACATCACTCTGAAATCTCTCAGACCAACATCTGGAGTAGTAATGCTCTCTGACA	1200
Db	79762	CATAGGAACATCACTCTGAAATCTCTCAGACCAACATCTGGAGTAGTAATGCTCTCTGACA	79821
Qy	1201	GTCTTAGAAGTTGCTACGCGCTGGATCTCCAAAGCGGTGACACACGCTGAGAGGAAT	1260
Db	79822	GTCTTAGAAGTTGCTACGCGCTGGATCTCCAAAGCGGTGACACACGCTGAGAGGAAT	79881
Qy	1261	GAGAAAGCTGGGCTCTCAGGTAAATCTGGCTTTTCAAAAGCCCCCTAATTTTACTGCA	1320
Db	79882	GAGAAAGCTGGGCTCTCAGGTAAATCTGGCTTTTCAAAAGCCCCCTAATTTTACTGCA	79941
Qy	1321	TAAATTAATTTGAATTCAGTAATTAATTTCTACAAATTTTCCATATAGTCACTACACAAT	1380
Db	79942	TAAATTAATTTGAATTCAGTAATTAATTTCTACAAATTTTCCATATAGTCACTACACAAT	80001
Qy	1381	ACCCCTCAGAGCAACTGGGCTTGTGTAATACATATCTATATATGAGACTGTGGCTCTT	1440
Db	80002	ACCCCTCAGAGCAACTGGGCTTGTGTAATACATATCTATATATGAGACTGTGGCTCTT	80061
Qy	1441	AAGGTAATGTTTATATACATAAGGCTCTTGGGCTTACATATAAAGGGGATTTGAGC	1500
Db	80062	AAGGTAATGTTTATATACATAAGGCTCTTGGGCTTACATATAAAGGGGATTTGAGC	80121
Qy	1501	AATGTGATACGAAGTCTTTTCTTCCACAGGCTCTCATATGTAAAGAAATTCATTAGATTGC	1560
Db	80122	AATGTGATACGAAGTCTTTTCTTCCACAGGCTCTCATATGTAAAGAAATTCATTAGATTGC	80181
Qy	1561	TGAAATATAGCTGATCTGTCAATTTCTGTCTCACTTATCATTAAGAAAGCATTAGAGTAG	1620
Db	80182	TGAAATATAGCTGATCTGTCAATTTCTGTCTCACTTATCATTAAGAAAGCATTAGAGTAG	80241

Qy	1621	GAACAAAACACGACATCTATGTATTTAGGAAGAACACAGCTGGTTTGCTCAATATAAAAT	1680
Db	80242	GAACAAAACACACATCTATGTATTTAGGAAGAACACAGCTGGTTTGCTCAATATAAAAT	80301
Qy	1681	AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGTCATAGGAATCTAT	1740
Db	80302	AAGAAAAAGAAACCATGTGAAAGTCAAAATATTTGTTTAATCAGTCATAGGAATCTAT	80366
Qy	1741	TAAAGATTTTGAATTCCTTATGATGAGACTATCTTGACTCAAGTGACAGTGTGAG	1800
Db	80362	TAAAGATTTTGAATTCCTTATGATGAGACTATCTTGACTCAAGTGACAGTGTGAG	80422
Qy	1801	CTTTTGGCCGTGTCCTCCACGTAGAAAGGAGCTTGTCAATAAAGCTTATATGTAC	1860
Db	80422	CTTTTGGCCGTGTCCTCCACGTAGAAAGGAGCTTGTCAATAAAGCTTATATGTAC	80481
Qy	1861	AGGTGCCAAGTTAAGTGCCTCCCAAGCTTCTCTTAAAGCATACTGAAATTTGTTTAGACT	1920
Db	80482	AGGTGCCAAGTTAAGTGCCTCCCAAGCTTCTCTTAAAGCATACTGAAATTTGTTTAGACT	80544
Qy	1921	TTTATGTGAAGTGAAGGAGATTAACAATCCCTCTGGGAGAACTTCTCCTCATCTTGGT	1980
Db	80542	TTTATGTGAAGTGAAGGAGATTAACAATCCCTCTGGGAGAACTTCTCCTCATCTTGGT	80601
Qy	1981	GAACTCATCTGCCAGAAATTC	2001
Db	80602	GAACTCATCTGCCAGAAATTC	80622

### RESULT 3

LOCUS	AX033914	18443 bp	DNA	linear	PAT 21-SEP-2000
DEFINITION	Sequence 6 from Patent WO951790.				
ACCESSION	AX033914				
VERSION	AX033914.1	GI:10280482			

SOURCE	unidentified
ORGANISM	unidentified
	unclassified sequences

REFERENCE	1
AUTHORS	Cancilla, M.R., Choo, K.H. and Du, S.D.
TITLE	A novel nucleic acid molecule
JOURNAL	Patent: WO 9851790-A 6 19-NOV-1998;

FEATURES  
CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART  
DESTEER DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)  
Location/qualifiers

```

/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

```

## ORIGIN

Query Match	99.2%	Score 1984.8	DB 6	Length 18443
Best Local Similarity	99.8%	Pred. NO. 0		
Matches 1997	Conservative	1	Mismatches 3	Indels 1
				Gaps 1

Oy	1	ATCTTTATTTGTATGAGAAAAAGAAAGATGAAAAATATATTAACGTGATATAGT	60
Db	4392	ATCTTTATTTGTATGAGAAAAAGAAAGATGAAAAATATATTAACGTGATATAGT	4351
Oy	61	TCAGGA - CCATGAGATTGCAGATGACAGAAACATCAATTCAAAACCAACGTACGCAAAAGCA	119
Db	4352	TCAGAACCCCTTGATGATTCAGATGACAGAAACCTCAATTCAAACCAACGTACGCAAAAGCA	4411
Oy	120	AAATATATTTGCTCATGTACCTTCTCAACAGAGAGGACAGATGAAAGGGGCTTTGGGAA	179
Db	4412	AAATATATTTGCTCATGTACCTTCTCAACAGAGAGGACAGATGAAAGGGGCTTTGGGAA	4471
Oy	180	CAAGGATTTGTCTCAAAATCTAGAGATATCTAGATTAATGTCACGATGGTCACTTCC	239
Db	4472	CAAGGATTTGTCTCAAAATCTAGAGATATCTAGATTAATGTCACGATGGTCACTTCC	4531

QY	240	IGTCCCTAGGATGGGTGATGGAGATGGTATGAGTCTTAATGGAGGAAAGATCATATTAAG	299
Db	4352	TGTCCCTTAGGATGGTGGTATGGCAGATGGTATGAGTCTTATATGGAGGAAAGATCATATTAAG	4599
QY	300	ATGAAGGTAGGAGCTAAGCAAAACAAGGGCAAGGGCCACTATATATCATGTAAAAATGGTTTT	359
Db	4592	ATGAAGGTAGGAGCTAAGCAAAACAAGGGCAAGGGCCACTATATATCATGTAAAAATGGTTTT	4655
QY	360	TTTGTGATGCTTCCCTTAATTTTCAAAAATGCTTCCAAACAATAATAGACACAGAAAATAAA	419
Db	4652	TTTGTGATGCTTCCCTTAATTTTCAAAAATGCTTCCAAACAAGTATAGACACAGAAAATAAA	4711
QY	420	CATAGGAGCTCTACTGATGGGTGGCTTTATCTTAAGCCTTGACTTGTCTTTTCAAGCTT	479
Db	4712	CATAGGAGCTCTACTGATGGGTGGCTTTATCTTAAGCCTTGACTTGTCTTTTCAAGCTT	4772
QY	480	ACTCATGCTCTGTATCTATGAGGCATATGCGCTGTAAAGCTTCTGACAGGTTTCTACTAA	539
Db	4772	ACTCATGCTCTGTATCTATGAGGCATATGCGCTGTAAAGCTTCTGACAGGTTTCTACTAA	4833
QY	540	GCTGGGTTCTCTAATATGAGCTCTATCCCATTTTCTGTGCTCACTCATGTGATCTTTCTCT	599
Db	4832	GCTGGGTTCTCTAATATGAGCTCTATCCCATTTTCTGTGCTCACTCATGTGATCTTTCTCT	4899
QY	600	TTTCTCACTCTGAGGACTGATGAGCTGTTTGTATGAGCTGCTTAAGCTTTGTGGGTT	659
Db	4892	TTTCTCACTCTGAGGACTGATGAGCTGTTTGTATGAGCTGCTTAAGCTTTGTGGGTT	4955
QY	660	TTTTTCTGGGAGACAATGCTTCAAGATTATCTTAAGCAAAATAAATAACAGACATGGGCC	719
Db	4952	TTTTTCTGGGAGACAATGCTTCAAGATTATCTTAAGCAAAATAAATAACAGACATGGGCC	5011

QY	720	AGGCGCTTCCCTCCACCTGGACCATGTTCCGAGGGCTCTTCACTAGTTTAGGTCA	779
Db	5012	AGGCTCTTCCCTCCACCTGACCATGTTCCGAGGGCTCTTCACTAGTTTAGGTCAA	5071
QY	780	GCATTCTTGGCAAAAGAGGCTTAGTTAAACATAGACATTCTAGCAATTATCTTTT	839
Db	5072	GCATTCTTGGCAAAAGAGGCTTAGTTAAACATAGACATTCTAGCAATTATCTTTT	5133
QY	840	GACATGTTGTAGATCTATTCACTTTTGTATTAAGCAATTCCTTAGTGAACCAACA	899
Db	5132	GACATGTTGTAGATCTATTCACTTTTGTATTAAGCAATTCCTTAGTGAACCAACA	5191
QY	900	CGAATCTAGCGCTCCGAGGATGAGAGGGTGGCTCTCAATACAGAGGATGTTCTGAGAGC	959
Db	5192	CGAATCTAGCGCTCTCGAGGATGAGAGGGTGGCTCTCAATACAGAGGATGTTCTGAGAGC	5251
QY	960	TGTATTTTGGGCACCTTAACCTATTCTCCACTTATAGGGCAGACAGACTGAATTAACACA	1019
Db	5252	TGTATTTTGGGCACCTTAACCTATTCTCCACTTATAGGGCAGACACTGAATTAACACA	5311
QY	1020	CTTAAGTTTGTATGTCATGTCATGTAGTTAGTCTCAGGCAGTGCAGGCTCAGAGGTGAACCTGA	1079
Db	5312	CTAAGTTTGTATGTCATGTCATGTAGTTAGTCTCAGGCAGTGCAGGCTCAGAGGTGAACCTGA	5371
QY	1080	CCTCTTAATGTGTGTCGACGCTTTCTTCCCTTCAGAAAGTCAGGTGTGTTTCTGCGACCTCT	1139
Db	5372	CCTCTTAATGTGTGTCGACGCTTTCTTCCCTTCAGAAAGTCAGGTGTGTTTCTGCGACCTCT	5431
QY	1140	CCATAGGAACATCAGTCTCTGAATCCTCGACACACATCTGAGATGATAGTCTCTGAC	1199
Db	5432	CCATAGGAACATCAGTCTCTGAATCCTCGACACACATCTGAGATGATAGTCTCTGAC	5491
QY	1200	AGTCTTAGAAGTTGTCTTACCGCTTGATCTCTCAAAGCGTGTGACACACCTGTAGAGAGAA	1258
Db	5492	AGTCTTAGAAGTTGTCTTACCGCTTGATCTCTCAAAGCGTGTGACACACCTGTAGAGAGAA	5551
QY	1260	TGAGAAAGCTGGGCTCTCAGGTAAATCTGTCTTTTTCACAAGGCCCTCTAATTTTACCTGC	1319
Db	5552	TGAGAAAGCTGGGCTCTCAGGTAAATCTGTCTTTTTCACAAGGCCCTCTAATTTTACCTGC	5611
QY	1320	ATAATTAATTTTGAATCACTGATTAATTTCTACAAATTTTCCATAGTATCTTACACACAA	1379







Db 37640 GACATGTTGTAACATCTATTCACATTTTGTATTTAAAGCAATTCCTTATGGAACCAACA 37699  
QY 900 CGAATTAAGCTGCTCTGGAATCAGAGGTGGCTCTCATATACAGATGTTCTAGAGAC 959  
Db 37700 CGAATTAAGCTGCTCTGGAATCAGAGGTGGCTCTCATATACAGATGTTCTAGAGAC 37759  
QY 960 TGTATTTTGGGCACTTAATCTTCTCACTATTAGGGCAGACACTGAAATTAACACA 1019  
Db 37760 TGTATTTTGGGCACTTAATCTTCTCACTATTAGGGCAGACACTGAAATTAACACA 37819  
QY 1020 CTATAGTTTGTCACTGATCTAGTATGTTCTCAGAGCTGAGGCTCAGAGTGTGAACCTGA 1079  
Db 37820 CTATAGTTTGTCACTGATCTAGTATGTTCTCAGAGCTGAGGCTCAGAGTGTGAACCTGA 37879  
QY 1080 CCTCTATGTGTCCAGCTTCTTCTCTCAGAAAGCAGCTGTGTTTCTGCTGACTCT 1139  
Db 37880 CCTCTATGTGTCCAGCTTCTTCTCTCAGAAAGCAGCTGTGTTTCTGCTGACTCT 37939  
QY 1140 CCATAGGAACATCATGCTGTAATCTCTCAACACCATCTGAGTATGTAAGTGTCTCTGAC 1199  
Db 37940 CCATAGGAACATCATGCTGTAATCTCTCAACACCATCTGAGTATGTAAGTGTCTCTGAC 37999  
QY 1200 AGTCCTAGAAGTGTCTACCGCTGGAATCTCCAAAGGCTGTGACACACCGGAGAGAGAA 1259  
Db 38000 AGTCCTAGAAGTGTCTACCGCTGGAATCTCCAAAGGCTGTGACACACCGGAGAGAGAA 38059  
QY 1260 TGAGAAAGCTGGGCTCTTCAGTAAATCTGCTTTTTCAGAAAGCCCTTAATTTTACTGC 1319  
Db 38060 TGAGAAAGCTGGGCTCTTCAGTAAATCTGCTTTTTCAGAAAGCCCTTAATTTTACTGC 38119  
QY 1320 ATATATATTTTGAATTCATGATTAATTTCTACAAATTTTCCATTAAGTCACTACACAA 1379  
Db 38120 ATATATATTTTGAATTCATGATTAATTTCTACAAATTTTCCATTAAGTCACTACACAA 38179  
QY 1380 TACCTCTCATGCAACCTTGGCTTGTCTAATACATATCTATATGAGAGCTGTCTCT 1439  
Db 38180 TACCTCTCATGCAACCTTGGCTTGTCTAATACATATCTATATGAGAGCTGTCTCTCT 38239  
QY 1440 TAAAGCGTAATGTTTATATGCACTAAGGCTCTTGGCTTACATATAAAGGGTATTTGAG 1499  
Db 38240 TAAAGCGTAATGTTTATATGCACTAAGGCTCTTGGCTTACATATAAAGGGTATTTGAG 38299  
QY 1500 CAATGTGATACAGAAAGCTTTTCTCCAGAGCTCATATGTAAAGATTCATTAGATTG 1559  
Db 38300 CAATGTGATACAGAAAGCTTTTCTCCAGAGCTCATATGTAAAGATTCATTAGATTG 38359  
QY 1560 CTGAAATATGATCTGATCTGCTCATTTCTGCTCACTTATCATTAAGAAAGCTATTAGCTTA 1619  
Db 38360 CTGAAATATGATCTGATCTGCTCATTTCTGCTCACTTATCATTAAGAAAGCTATTAGCTTA 38419  
QY 1620 GGAACAAAACCTCAATCTATGTAATAGAAAGCAAGCTGTGTTTCTCAATATAAAA 1679  
Db 38420 GGAACAAAACCTCAATCTATGTAATAGAAAGCAAGCTGTGTTTCTCAATATAAAA 38479  
QY 1680 TAAAGAAAAGAAAACCATGTGAAAGTCAAAATATTTTGTATCAAGCTCATTTGAGAACTTA 1739  
Db 38480 TAAAGAAAAGAAAACCATGTGAAAGTCAAAATATTTTGTATCAAGCTCATTTGAGAACTTA 38539  
QY 1740 TTTAAAGATATTTGAATTTCTTTATGATGAGAACTATCTTCAAGTGAACAGTGTGA 1799  
Db 38540 TTTAAAGATATTTGAATTTCTTTATGATGAGAACTATCTTCAAGTGAACAGTGTGA 38599  
QY 1800 GCTTTTGGCTGTGGCTCCCAAGTATAGAAAGAGGCTTTGTCAATAAAGCTTTATATGTA 1859  
Db 38600 GCTTTTGGCTGTGGCTCCCAAGTATAGAAAGAGGCTTTGTCAATAAAGCTTTATATGTA 38659  
QY 1860 CAGGTGCAAGTAAAGTGCACCAAGCTTCTTTAAAGCACTATGATATTTGTTTTAGAC 1919  
Db 38660 CAGGTGCAAGTAAAGTGCACCAAGCTTCTTTAAAGCACTATGATATTTGTTTTAGAC 38719  
QY 1920 TTTTATGTAACGTAGAGGAATTAACAAATCCCTCTGGGAGAACTTCTCTCATCTTGG 1979

Db 38720 TTTTATGTAACGTAGAGGAATTAACAAATCCCTCTGGGAGAACTTCTCTCATCTTGG 38779  
QY 1980 TGAAGTCAATTTCTGCCAGAAATTC 2001  
Db 38780 TGAAGTCAATTTCTGCCAGAAATTC 38801  
RESULT 5  
AX033939 39238 bp DNA linear PAT 21-SEP-2000  
LOCUS AX033939  
DEFINITION Sequence 31 from Patent WO9851790.  
ACCESSION AX033939  
VERSION AX033939.1 GI:10280508  
KEYWORDS  
SOURCE unidentifed  
ORGANISM unidentifed  
REFERENCE 1 unclassified sequences.  
AUTHORS Cancelli,M.R., Choo,K.H. and Du,S.D.  
TITLE A novel nucleic acid molecule  
JOURNAL Patent: WO 9851790-A 31 19-NOV-1998;  
CANCELLI MICHAEL, ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART  
DESIRE DU (AU) ; AMRAD OPERATIONS PRY LTD (AU)  
FEATURES  
source location/Qualifiers  
1..39238  
/organism="unidentifed"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 94.7%; Score 1895.2; DB 6; Length 39238;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1908; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
QY 1 ATCTTATTTGTATGAGAAAAAGAGAGTGAATAATATTTAAGTGCATATAGT 60  
Db 37327 ATCTTATTTGTATGAGAAAAAGAGAGTGAATAATATTTAAGTGCATATAGT 37386  
QY 61 TCAGGA-CCATGATTCAGAAAGTGAAGAACTCAATTCAAACCAAGTAAAGAA 119  
Db 37387 TCAGGAACCTTGATTCAGAAAGTGAAGAACTCAATTCAAACCAAGTAAAGAA 37446  
QY 120 AATATATTTGGCTCATGTAACCTTCTCACAGAGAGGCGAGATGGAAGGGCTTTGGGA 179  
Db 37447 AATATATTTGGCTCATGTAACCTTCTCACAGAGAGGCGAGATGGAAGGGCTTTGGGA 37506  
QY 180 CAAGAAATTTGTTCCAAATTTCTAGAAATATCTAGAGTTAGTCAGAGATGGTCACTTCC 239  
Db 37507 CAAGAAATTTGTTCCAAATTTCTAGAAATATCTAGAGTTAGTCAGAGATGGTCACTTCC 37566  
QY 240 TGTCCCTGAGTGTGTGATGCGATGGTATGAGTCTTATGGAGAAAGAGTGCATGTAG 299  
Db 37567 TGTCCCTGAGTGTGTGATGCGATGGTATGAGTCTTATGGAGAAAGAGTGCATGTAG 37626  
QY 300 ATGAAGTATGGGCTTAAGCAACAGAGGCGCACTATATCATGCTTAAATATGTTTT 359  
Db 37627 ATGAAGTATGGGCTTAAGCAACAGAGGCGCACTATATCATGCTTAAATATGTTTT 37686  
QY 360 TTTTGAATGTCCTTCAATTTTCAAAATGCTTCAACAAAGTATGAGCAACAGAAAAAGA 419  
Db 37687 TTTTGAATGTCCTTCAATTTTCAAAATGCTTCAACAAAGTATGAGCAACAGAAAAAGA 37746  
QY 420 CATAGGAATCTACTAGTGTGGTCTTTTATCTTAAAGCTTGTATCTTTTCAAGCTT 479  
Db 37747 CATAGGAATCTACTAGTGTGGTCTTTTATCTTAAAGCTTGTATCTTTTCAAGCTT 37806  
QY 480 ACTACTGCTTGTACTGAGGCGCATATGCTCTGTAAAGCTTCTGAGGGTTTCTACTAA 539  
Db 37807 ACTACTGCTTGTACTGAGGCGCATATGCTCTGTAAAGCTTCTGAGGGTTTCTACTAA 37866  
QY 540 GCTGGTTCCTTATATAGTCTCTCCCATTTCTGTTGCTCTCACTAGATCTTCTCT 599  
Db 37867 GCTGGTTCCTTATATAGTCTCTCTCCCATTTCTGTTGCTCTCACTAGATCTTCTCT 37926

QY 600 TTTCTCACCTCTGGAGCTGGTGGCTGTTGATGACCTGCTAGCTTTGCTTGGGTT 659  
 Db 37927 TTTCTCACCTCTGGAGCTGGTGGCTGTTGATGACCTGCTAGCTTTGCTTGGGTT 37986  
 QY 660 TTTTCTGGGAGACATGCTCTTCAAGATTATCTAGACCATAAATCTACAGCCCTGGGCC 719  
 Db 37987 TTTTCTGGGAGACATGCTCTTCAAGATTATCTAGACCATAAATCTACAGCCCTGGGCC 38046  
 QY 720 AGGCTCTTCTCTCTCCACCTGACCATGTTCCAGAGGCTCTTCACTTATGTTAGTCAA 779  
 Db 38047 AGGCTCTTCTCTCTCCACCTGACCATGTTCCAGAGGCTCTTCACTTATGTTAGTCAA 38106  
 QY 780 GCATCTTGGCAAAAGAAAGGCTAGTTAAACAATAGACATTTAGCAATGATTTCTTTT 839  
 Db 38107 GCATCTTGGCAAAAGAAAGGCTAGTTAAACAATAGACATTTAGCAATGATTTCTTTT 38166  
 QY 840 GACATGTTGTAAGTCTATTCACATTTTGTAAATTAAGCATTCCTTATGAAACACACA 899  
 Db 38167 GACATGTTGTAAGTCTATTCACATTTTGTAAATTAAGCATTCCTTATGAAACACACA 38226  
 QY 900 CGAATCTAGCTGCTCTGGAATGAGGCTGCTCTCAATACAGAGTGTCTTAGAGAGC 959  
 Db 38227 CGAATCTAGCTGCTCTGGAATGAGGCTGCTCTCAATACAGAGTGTCTTAGAGAGC 38286  
 QY 960 TGTATTTTGGGCACTTAATCTATCTCCACTTAAGGCGACAGACTGAAATTAACACA 1019  
 Db 38287 TGTATTTTGGGCACTTAATCTATCTCCACTTAAGGCGACAGACTGAAATTAACACA 38346  
 QY 1020 CTAAGTTTGTATGCTCATGCTAGTTAGTCTCAGGAGTGCAGCTCAGAGTGAATCTGA 1079  
 Db 38347 CTAAGTTTGTATGCTCATGCTAGTTAGTCTCAGGAGTGCAGCTCAGAGTGAATCTGA 38406  
 QY 1080 CCTCTTATGTTGTCCAGCCCTTTCTTCTTCAAGAGTCACTGTGTTTCTGCTGACT 1139  
 Db 38407 CCTCTTATGTTGTCCAGCCCTTTCTTCTTCAAGAGTCACTGTGTTTCTGCTGACT 38466  
 QY 1140 CCATAGGAACATGATCTGTAATCTCTACACACCATCTGAGTATGATGCTCTGAC 1199  
 Db 38467 CCATAGGAACATGATCTGTAATCTCTACACACCATCTGAGTATGATGCTCTGAC 38526  
 QY 1200 AGTCTAGAGTGTCTCACCCTGAGTCTCCAAAGGCTGACACACCGTAGAGAGAA 1259  
 Db 38527 AGTCTAGAGTGTCTCACCCTGAGTCTCCAAAGGCTGACACACCGTAGAGAGAA 38586  
 QY 1260 TGAAGAAAGCTGGGCTCTTCAAGTAAATCTTGTCTTTTCAAGACCCCTTAATTTACT 1319  
 Db 38587 TGAAGAAAGCTGGGCTCTTCAAGTAAATCTTGTCTTTTCAAGACCCCTTAATTTACT 38646  
 QY 1320 ATTAATTTTGAATCACTGATTAATTTCTACAAATTTCCCATTAAGTCACTACACAA 1379  
 Db 38647 ATTAATTTTGAATCACTGATTAATTTCTACAAATTTCCCATTAAGTCACTACACAA 38706  
 QY 1380 TACCTCTCATGCAACCTGGCTTGTCTTAATCATATCTAATATGAGAGCTGTGCTGT 1439  
 Db 38707 TACCTCTCATGCAACCTGGCTTGTCTTATCATATCTAATATGAGAGCTGTGCTGT 38766  
 QY 1440 TAAAGCTAAATGTTTTATATGACATAAGGCTTGTGCTTACATATAAAGGGTATTTGAG 1499  
 Db 38767 TAAAGCTAAATGTTTTATATGACATAAGGCTTGTGCTTACATATAAAGGGTATTTGAG 38826  
 QY 1500 CAATGATATACAGAGCTTTTCTCCAGAGCTCATATGTAAGAATTCATTAGATTG 1559  
 Db 38827 CAATGATATACAGAGCTTTTCTCCAGAGCTCATATGTAAGAATTCATTAGATTG 38886  
 QY 1560 CTGAATATAGACTGATCTGCTCATTTCTGCTCAGCTATCATAGGAAGTCAATAGCTAA 1619  
 Db 38887 CTGAATATAGACTGATCTGCTCATTTCTGCTCAGCTATCATAGGAAGTCAATAGCTAA 38946  
 QY 1620 GGAACAAAACATCAATCTATGTAATTAAGAAACAAGCTGTGTTTGTCAATATATAA 1679  
 Db 38947 GGAACAAAACATCAATCTATGTAATTAAGAAACAAGCTGTGTTTGTCAATATATAA 39006

QY 1680 TAAAGAAAAGAACCATGATGAAAGTCAAAATATTTGTTAATCAAGTCAATGAGATCTA 1739  
 Db 39007 TAAAGAAAAGAACCATGATGAAAGTCAAAATATTTGTTAATCAAGTCAATGAGATCTA 39066  
 QY 1740 TTAAGAAATTTGAATTTCTTATATGATGAGAACTATCTTACTCAAGTGAAGTGTGA 1799  
 Db 39067 TTAAGAAATTTGAATTTCTTATATGATGAGAACTATCTTACTCAAGTGAAGTGTGA 39126  
 QY 1800 GCTTTTGGGCTGTGGTCCCTACGTAAGAGAGGCTTTGCTATTAAGTCTTATATGTA 1859  
 Db 39127 GCTTTTGGGCTGTGGTCCCTACGTAAGAGAGGCTTTGCTATTAAGTCTTATATGTA 39186  
 QY 1860 CAGTGCCAAAGTTAAGTGCCCAAGCTGCTCTTAAGACATCTGATTTTG 1911  
 Db 39187 CAGTGCCAAAGTTAAGTGCCCAAGCTGCTCTTAAGACATCTGATTTTG 39238  
 RESULT 6  
 AF042484 80155 bp DNA linear pri 22-JUN-2000  
 LOCUS AF042484  
 DEFINITION Homo sapiens sequence of the 80 kb core centromere protein binding region of the mar del(10) neocentromere corresponding to human chromosome 10q25.2, complete sequence.  
 ACCESSION AF042484  
 VERSION AF042484.1 GI:4205782  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 80155)  
 AUTHORS Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.  
 TITLE Sequence analysis of an 80 kb human neocentromere  
 JOURNAL Hum. Mol. Genet. 8 (2), 217-227 (1999)  
 PUBMED 9931329  
 REFERENCE 2 (bases 1 to 80155)  
 AUTHORS Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,A.  
 TITLE Direct Submision  
 JOURNAL Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville, Melbourne, Victoria 3052, Australia  
 FEATURES  
 source  
 1..80155  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /map="10q25.2"  
 /note="Sequence information obtained from BAC clones prepared from cell line BE2C1-18-51"  
 complement(join(49..525,3628..4545,4849..5053))  
 /rpt\_family="LIMBc"  
 /rpt\_type=dispersed  
 join(560..597,3184..3281)  
 /rpt\_family="MER58"  
 /rpt\_type=dispersed  
 complement(join(598..2553,2716..3183))  
 /rpt\_family="TIGGR1"  
 /rpt\_type=dispersed  
 2558..2581  
 /standard\_name="T1"  
 /note="100% homology"  
 /rpt\_type=direct  
 /rpt\_unit="(cagcacagtg) 2"  
 2570..2715  
 /rpt\_family="AluSg/x"  
 /rpt\_type=dispersed  
 3021..3050  
 /standard\_name="T2"  
 /note="83.3% homology"  
 /rpt\_type=direct  
 /rpt\_unit="(taacaagtc) 3"

```
repeat_region 3283. .3558 /note="AluSx"
/rpt_family="AluSx"
/rpt_type=dispersed
repeat_region 4173. .4211 /note="82.1% homology"
/standard_name="T3"
/rpt_type=direct
/rpt_unit="(aaaaaataattc)3"
complement(4546. .4848)
/rpt_family="AluSx"
/rpt_type=dispersed
complement(5909. .6208)
/rpt_family="AluSg"
/rpt_type=dispersed
join(6232. .6573,6638. .6688)
/rpt_family="MLR2Cb"
/rpt_type=dispersed
6598. .6645
/standard_name="AC24"
/note="100% homology"
/rpt_type=tandem
/rpt_unit="ac"
6598. .6637
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="ca"
7091. .7134
/rpt_family="MIR"
/rpt_type=dispersed
7808. .7911
/rpt_family="L2"
/rpt_type=dispersed
8349. .8524
/rpt_family="MIR"
/rpt_type=dispersed
complement(8839. .9124)
/rpt_family="AluNb"
/rpt_type=dispersed
9180. .9280
/rpt_family="MIR"
/rpt_type=dispersed
10070. .10170
/note="low complexity"
/rpt_family="purine-rich"
10097. .10136
/standard_name="GA1"
/note="82.5% homology"
/rpt_type=direct
/rpt_unit="(gaaagaagg)4"
10175. .10339
/rpt_family="FRAM"
/rpt_type=dispersed
complement(10899. .11197)
/rpt_family="AluSx"
/rpt_type=dispersed
join(11627. .11757,12434. .12512,13443. .13562)
/note="98% identity"
/db_xref="dbEST:AA774571"
11996. .12028
/standard_name="AT4"
/note="100% homology"
/rpt_type=tandem
/rpt_unit="(attc)8"
complement(12010. .12312)
/rpt_family="AluNb"
/rpt_type=dispersed
12327. .12433
/rpt_family="L2"
/rpt_type=dispersed
complement(13946. .14235)
/rpt_family="AluSg"
/rpt_type=dispersed
15130. .15159

repeat_region /note="low complexity"
/rpt_family="AT-rich"
15164. .15753
/standard_name="AT28"
/note="82.1% homology"
/rpt_type=direct
/rpt_unit="(atgatcatatgctatagacataac)21"
join(15170. .15289,15410. .15612)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="ca"
15617. .15731
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="cata"
16004. .16100
/rpt_family="L2"
/rpt_type=dispersed
complement(16128. .16364)
/rpt_family="L2"
/rpt_type=dispersed
complement(16407. .16592)
/rpt_family="L2"
/rpt_type=dispersed
complement(16657. .16889)
/rpt_family="MIR"
/rpt_type=dispersed
16914. .16936
/note="low complexity"
/rpt_family="AT-rich"
17102. .17386
/rpt_family="AluSx"
/rpt_type=dispersed
17406. .17564
/standard_name="GA2"
/note="81.8% homology"
/rpt_type=direct
17416. .17543
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="gaaa"
17647. .17702
/rpt_family="MIR"
/rpt_type=dispersed
complement(19863. .20161)
/rpt_family="AluSx"
/rpt_type=dispersed
complement(20842. .20913)
/rpt_family="L2"
/rpt_type=dispersed
complement(21366. .21414)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="gaaa"
21375. .21639
/standard_name="CT1"
/note="75% homology"
/rpt_type=dispersed
complement(21415. .21536)
/note="microsatellite"
/rpt_type=tandem
/rpt_unit="cgaa"
22560. .22604
/rpt_family="LM4"
/rpt_type=dispersed
22609. .22646
/rpt_family="purine-rich"
22409. .22646
/rpt_family="LM4"
/rpt_type=dispersed
22560. .22604
/standard_name="T4"
/note="80% homology"
/rpt_type=direct
/rpt_unit="(atactacataatc)3"
```

Query Match 94.7%; Score 1895.2; DB 8; Length 80155;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1908; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 ATCTTATTTGATGAGAAAAAGAGAGTGAATAATATATTAACGTGATATAGT 60  
DB 78244 ATCTTATTTGATGAGAAAAAGAGAGTGAATAATATTAACGTGATATAGT 78303

QY 61 TCAGGA-CCATGATTTGCAAGTGAAGAACTCAATTCAACCAAGTAAAGTCAAAAGGA 119  
DB 78304 TCAGGA-CCATGATTTGCAAGTGAAGAACTCAATTCAACCAAGTAAAGTCAAAAGGA 78363

QY 120 AATATATTGGGCTCATGTAACTTTCACAGAGAGGAGAGTGAAGGGGCTTTGGGAA 179  
DB 78364 AATATATTGGGCTCATGTAACTTTCACAGAGAGGAGAGTGAAGGGGCTTTGGGAA 78423

QY 180 CAAGAGATTTGTTCTCAAAATTTAGGAATPACTAGATTAGTCCAGATGGGTCACTTCC 239  
DB 78424 CAAGAGATTTGTTCTCAAAATTTAGGAATPACTAGATTAGTCCAGATGGGTCACTTCC 78483

QY 240 TGTCCCTGAGGTGGTGGTGAAGTGTGAGTCTTATGGAGGAAAGTGCATGTTAGG 299  
DB 78484 TGTCCCTGAGGTGGTGGTGAAGTGTGAGTCTTATGGAGGAAAGTGCATGTTAGG 78543

QY 300 ATGAAGGTAGGGCTTAAGCAAAAGAGGCAAGGGCACTATATGATGCTAAATGGTTTT 359  
DB 78544 ATGAAGGTAGGGCTTAAGCAAAAGAGGCAAGGGCACTATATGATGCTAAATGGTTTT 78603

QY 360 TTTTGAATGTTCTTCTTAATTTCAAAATGCTTCCAAACAAAGTGAACACAGAGAAAAAGAA 419  
DB 78604 TTTTGAATGTTCTTCTTAATTTCAAAATGCTTCCAAACAAAGTGAACACAGAGAAAAAGAA 78663

QY 420 CATAGGACTCTAATGAGGTGGTCTTTTATCTTAAGCTTTGATCTTTTCAAGCTT 479  
DB 78664 CATAGGACTCTAATGAGGTGGTCTTTTATCTTAAGCTTTGATCTTTTCAAGCTT 78723

QY 480 ACTCACTGCTTGAACCTGAGGCGCATATGCCCTGTAAGGTTTGAAGGTTTCTACTAA 539  
DB 78724 ACTCACTGCTTGAACCTGAGGCGCATATGCCCTGTAAGGTTTGAAGGTTTCTACTAA 78783

QY 540 GCTGAGTCTTATATGAGTCTCTCCCATTTCTGTTGCTCACTAGTGAATCTTCTGT 599  
DB 78784 GCTGAGTCTTATATGAGTCTCTCTCCCATTTCTGTTGCTCACTAGTGAATCTTCTGT 78843

QY 600 TTTTCTCACTCTGAGACTGGTGGTCTTTTGAATGAGCTGCTTGAAGCTTTTGGGT 659  
DB 78844 TTTTCTCACTCTGAGACTGGTGGTCTTTTGAATGAGCTGCTTGAAGCTTTTGGGT 78903

QY 660 TTTTCTGGGGAACAATGCTTGAAGTTATCTTGAACCAAAATAACTACAGCCCTGGGC 719  
DB 78904 TTTTCTGGGGAACAATGCTTGAAGTTATCTTGAACCAAAATAACTACAGCCCTGGGC 78963

QY 720 AAGCTCTTCTCTCCCAACTGAGCAATGTTCCAGAGGCTTCTCACTTATGTTAGTCA 779  
DB 78964 AAGCTCTTCTCTCTCCCAACTGAGCAATGTTCCAGAGGCTTCTCACTTATGTTAGTCA 79023

QY 780 GCATTTCTGGCAAAAGAAAGGCTAGTTAAACAATAGCAATTCAGCAATTTGATTTCTTTT 839  
DB 79024 GCATTTCTGGCAAAAGAAAGGCTAGTTAAACAATAGCAATTCAGCAATTTGATTTCTTTT 79083

QY 840 GACATGTTGTAAGATCTATTCACATTTTGTATTAAGCATTTCCCTATGAGAAACAACA 899  
DB 79084 GACATGTTGTAAGATCTATTCACATTTTGTATTAAGCATTTCCCTATGAGAAACAACA 79143

QY 900 CGAATTAAGCTGCTCTGCAATGAGAGTGGCTCTCTCAATACAGATGTTCTAGAGAGC 959  
DB 79144 CGAATTAAGCTGCTCTGCAATGAGAGTGGCTCTCTCAATACAGATGTTCTAGAGAGC 79203

QY 960 TGTATTTTGGGCACTTAATCTTCTCACTAATTAGGGCAAGCACTGAATTAACAACA 1019  
DB 79204 TGTATTTTGGGCACTTAATCTTCTCACTAATTAGGGCAAGCACTGAATTAACAACA 79263

QY 1020 CTAACTTGTGATGCTCATGATGATGATGCTGACGAGAGTGCAGCTCAGAGTGAACCTGA 1079  
DB 79264 CTAACTTGTGATGCTCATGATGATGATGCTGACGAGAGTGCAGCTCAGAGTGAACCTGA 79323

QY 1080 CCTCTTATGTTGTCAGGCTTTTCTTCTTCAAGAGTCAAGCTGTGTTTCTGCTGACTCT 1139  
DB 79324 CCTCTTATGTTGTCAGGCTTTTCTTCTTCAAGAGTCAAGCTGTGTTTCTGCTGACTCT 79383

QY 1140 CCATAGGAACATCAGCTCGAATCCGACAGACCAACATCTGGAGTGAATGCTCCTGAC 1199  
DB 79384 CCATAGGAACATCAGCTCGAATCCGACAGACCAACATCTGGAGTGAATGCTCCTGAC 79443

QY 1200 AGTCTAGAAATGTTCTACCGCTGATCTCCAAAGCTGTGAACAACCTGAGAGAGAA 1259  
DB 79444 AGTCTAGAAATGTTCTACCGCTGATCTCCAAAGCTGTGAACAACCTGAGAGAGAA 79503

QY 1260 TGAAGAAAGCTGGCTCTTCAAGGTAAATCTTGCTTTTCAAGAGCCCTTAATTTTACTGC 1319  
DB 79504 TGAAGAAAGCTGGCTCTTCAAGGTAAATCTTGCTTTTCAAGAGCCCTTAATTTTACTGC 79563

QY 1320 ATPATATTTTGAATTCATCTGATTAATTTCTACAAATTTCCCAATGATCTACACACA 1379  
DB 79564 ATPATATTTTGAATTCATCTGATTAATTTCTACAAATTTCCCAATGATCTACACACA 79623

QY 1380 TACCTCTCATGCAACACTTGCTTGTGTAATACATATCTAATGAGAGCTGTGCTCT 1439  
DB 79624 TACCTCTCATGCAACACTTGCTTGTGTAATACATATCTAATGAGAGCTGTGCTCT 79683

QY 1440 TAAAGGTAAATGTTTATATGACATTAAGGCTTGGCTTACATATTAAGGGGTATTGAG 1499  
DB 79684 TAAAGGTAAATGTTTATATGACATTAAGGCTTGGCTTACATATTAAGGGGTATTGAG 79743

QY 1500 CAATGATTAACGAAGTCTTTCTCAAGAGTCTCATATGTAAGAATTCATTAAGTTGG 1559  
DB 79744 CAATGATTAACGAAGTCTTTCTCAAGAGTCTCATATGTAAGAATTCATTAAGTTGG 79803

QY 1560 CTGAATATAGCTGATCTGTCATTTCTGCTCACTTATCATTAAGAGTCAATTAAGCTAA 1619  
DB 79804 CTGAATATAGCTGATCTGTCATTTCTGCTCACTTATCATTAAGAGTCAATTAAGCTAA 79863

QY 1620 GGAACAAAACATCAATCTATGTAATTAAGAAACAAGCTGTTTGTCAATATTAATA 1679  
DB 79864 GGAACAAAACATCAATCTATGTAATTAAGAAACAAGCTGTTTGTCAATATTAATA 79923

QY 1680 TAAAGAAAAAGAAACCATGTAAGTAAGTAATTTGTTTAATCAGGTCAATGAATCTA 1739  
DB 79924 TAAAGAAAAAGAAACCATGTAAGTAAGTAATTTGTTTAATCAGGTCAATGAATCTA 79983

QY 1740 TTAATAAGTATTTGAATCTTATGATGAGACTATCTTCAAGTCAAGTGAAGTGA 1799  
DB 79984 TTAATAAGTATTTGAATCTTATGATGAGACTATCTTCAAGTCAAGTGAAGTGA 80043

QY 1800 GCTTTTGGCTGTGGTCCCTACGTAAGAAAGAGGCTTTGTATTAAGTCTTATATGTA 1859  
DB 80044 GCTTTTGGCTGTGGTCCCTACGTAAGAAAGAGGCTTTGTATTAAGTCTTATATGTA 80103

QY 1860 CAGGTGCCAAGTTAAGTGGCCAAAGCTGCTTTAAAGACATCTGATTTTG 1911  
DB 80104 CAGGTGCCAAGTTAAGTGGCCAAAGCTGCTTTAAAGACATCTGATTTTG 80155

RESULT 7  
AF222854 80202 bp DNA linear PRI 17-JUL-2000  
LOCUS AF222854  
DEFINITION Homo sapiens clone NC chromosome 10 map 10q25.2 genomic sequence.  
ACCESSION AF222854  
VERSION AF222854.1 GI:9246844  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.

REFERENCE	1 (bases 1 to 80202)
AUTHORS	Barry,A.E., Bateman,M., Howman,E.V., Cancellia,M.R., Tainton,K.M., Irvine,D.V., Saffery,R. and Choo,K.H.
TITLE	The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification
JOURNAL	Genome Res. 10 (6), 832-838 (2000)
PUBMED	10854414
REFERENCE	2 (bases 1 to 80202)
AUTHORS	Barry,A.E.
TITLE	Direct Submission
JOURNAL	Submitted (11-JAN-2000) Chromosome Reseach Unit, The Murdoch Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd., Parkville,, Melbourne, Victoria 3052, Australia
REMARK	Human genomic sequence from 10q25.2, clone1p=NC, second release
FEATURES	Location/Qualifiers
SOURCE	1. 80202
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="10"
	/map="10q25.2"
	/clone="NC"
repeat_region	complement(598..3183)
	/rpt_family="Tigert1"
repeat_region	2587..2715
	/rpt_family="Alu"
repeat_region	3283..3553
	/rpt_family="Alu"
repeat_region	complement(4548..4809)
	/rpt_family="Alu"
repeat_region	complement(5919..6211)
	/rpt_family="Alu"
repeat_region	6257..6599
	/rpt_family="MT2B3"
repeat_region	complement(6848..9123)
	/rpt_family="Alu"
repeat_region	complement(10913..11198)
	/rpt_family="Alu"
repeat_region	complement(12062..12313)
	/rpt_family="Alu"
repeat_region	complement(13957..14256)
	/rpt_family="Alu"
repeat_region	17133..17424
	/rpt_family="Alu"
repeat_region	complement(19882..20197)
	/rpt_family="Alu"
repeat_region	complement(22664..22956)
	/rpt_family="Alu"
repeat_region	complement(23935..24113)
	/rpt_family="MIR"
repeat_region	24585..24884
	/rpt_family="Alu"
repeat_region	complement(25618..25973)
	/rpt_family="THER1"
repeat_region	26732..26970
	/rpt_family="MER2"
repeat_region	27173..27449
	/rpt_family="Alu"
repeat_region	27526..27796
	/rpt_family="Alu"
repeat_region	28702..28978
	/rpt_family="Alu"
repeat_region	complement(30668..30736)
	/rpt_family="MLT1"
repeat_region	complement(31460..31993)
	/rpt_family="MER1"
repeat_region	complement(32130..32248)
	/rpt_family="MLT1"
repeat_region	32685..32961
	/rpt_family="Alu"
repeat_region	33113..33403
	/rpt_family="Alu"

[illegible]



QY 120 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGATGAGAGGGCTTTGGGAA 179  
 |||||  
 Db 78411 AATATATTGGCTCATGTAACCTTCTCACAGAGAGGGCAGATGAGAGGGCTTTGGGAA 78470  
 |||||  
 QY 180 CAAGAGAAATTTCTCAAAATTTCTAGAGATATAGAGATTTAGTCCAGAGATGGGTCACTTCC 239  
 |||||  
 Db 78471 CAAGAGAAATTTCTCAAAATTTCTAGAGATTTAGTCCAGAGATGGGTCACTTCC 78530  
 |||||  
 QY 240 TGTCCCTGAGGTGTGTAGCCGATGTAGAGTCTTATGTGAGAGAAAGATGATGATAGG 239  
 |||||  
 Db 78531 TGTCCCTGAGGTGTGTAGCCGATGTAGAGTCTTATGTGAGAGAAAGATGATAGG 78590  
 |||||  
 QY 300 ATGAAGTAGAGGCTTAAGCAAAACAAGGCAAGGGCCACTATATCATGTCTAAATGGTTTT 359  
 |||||  
 Db 78591 ATGAAGTAGAGGCTTAAGCAAAACAAGGCAAGGGCCACTATATCATGTCTAAATGGTTTT 78650  
 |||||  
 QY 360 TTTTGAATGCTTCTTAATTTCACAAATGCTTCCAAACAAGTAGCACACAGAAAAAGAA 419  
 |||||  
 Db 78651 TTTTGAATGCTTCTTAATTTCACAAATGCTTCCAAACAAGTAGCACACAGAAAAAGAA 78710  
 |||||  
 QY 420 CATAGGAGACTCTATGCTGGGTGCTTTTATCTTAAGCTTTGATCTTGTCTTTTCAAGCTT 479  
 |||||  
 Db 78711 CATAGGAGACTCTATGCTGGGTGCTTTTATCTTAAGCTTTGATCTTGTCTTTTCAAGCTT 78770  
 |||||  
 QY 480 ACTCATGCTGTATACCGGAGGCAATAGCCCTGTAAAGGCTTGTGAGGGTTTCTACTAA 539  
 |||||  
 Db 78771 ACTCATGCTGTATACCGGAGGCAATAGCCCTGTAAAGGCTTGTGAGGGTTTCTACTAA 78830  
 |||||  
 QY 540 GCTGGGTCTCTTATATGCTGTCTCTCCATTTCTGTGCTCTCATCTAGTATGATCTTCTCT 599  
 |||||  
 Db 78831 GCTGGGTCTCTTATATGCTGTCTCTCCATTTCTGTGCTCTCATCTAGTATGATCTTCTCT 78890  
 |||||  
 QY 600 TTTTCTCACTCTGGGACCTGTGTGCTTTTGTATGACATGCTCTTACCTTTGCTTTGGGTT 659  
 |||||  
 Db 78891 TTTTCTCACTCTGGGACCTGTGTGCTTTTGTATGACATGCTCTTACCTTTGCTTTGGGTT 78950  
 |||||  
 QY 660 TTTTCTGGGGACAATGCTCTTCAAGATTATTCCTAGAACAAATTAACATACAGCCCTGGGCT 719  
 |||||  
 Db 78951 TTTTCTGGGGACAATGCTCTTCAAGATTATTCCTAGAACAAATTAACATACAGCCCTGGGCT 79010  
 |||||  
 QY 720 AGGCTTCTCTCTCTCCAACTGACACATGTTCCCAAGGCTCTTCAACCTTATGTTAGGTCAA 779  
 |||||  
 Db 79011 AGGCTTCTCTCTCTCCAACTGACACATGTTCCCAAGGCTCTTCAACCTTATGTTAGGTCAA 79070  
 |||||  
 QY 780 GCATTTTGGCAAAAGAAAGGCTTATTAACAATTAACAATTTAGCAATTTGATTTCTTTT 839  
 |||||  
 Db 79071 GCATTTTGGCAAAAGAAAGGCTTATTAACAATTAACAATTTGATTTCTTTT 79130  
 |||||  
 QY 840 GACATGTGTAAAGATCTATTTCACAATTTGTAATTAAGATTCCTTATGAGAAACCAACA 899  
 |||||  
 Db 79131 GACATGTGTAAAGATCTATTTCACAATTTGTAATTAAGATTCCTTATGAGAAACCAACA 79190  
 |||||  
 QY 900 CGAAGTAAGCTGCTCTCTGGAATGACAGGTGGCTCTCTCAATACAGATGTTCTAGAGAC 959  
 |||||  
 Db 79191 CGAAGTAAGCTGCTCTCTGGAATGACAGGTGGCTCTCTCAATACAGATGTTCTAGAGAC 79250  
 |||||  
 QY 960 TGTATTTTGGGCACTTAATCTATTTCTCACTACTTAAAGGCAAGCACTGAATTAACACA 1019  
 |||||  
 Db 79251 TGTATTTTGGGCACTTAATCTATTTCTCACTACTTAAAGGCAAGCACTGAATTAACACA 79310  
 |||||  
 QY 1020 CTAAAGTTTGTCAATGCTCAATGTTATGTTCTCAGGAGGTGACAGCTTCAGAGTGAACCTAA 1079  
 |||||  
 Db 79311 CTAAAGTTTGTCAATGCTCAATGTTATGTTCTCAGGAGGTGACAGCTTCAGAGTGAACCTAA 79370  
 |||||  
 QY 1080 CCTCTATGTTGTCTCAGGCTTCTCTCTTCAAGAGTCAAGCTGTGTTTCTGCTGACTCT 1139  
 |||||  
 Db 79371 CCTCTATGTTGTCTCAGGCTTCTCTCTTCAAGAGTCAAGCTGTGTTTCTGCTGACTCT 79430  
 |||||  
 QY 1140 CCATAGGAACATAGTCTGTAATCTCTCAGACCACTCTGAGTATGTAAGTGTCTCTGAC 1199  
 |||||  
 Db 79431 CCATAGGAACATAGTCTGTAATCTCTCAGACCACTCTGAGTATGTAAGTGTCTCTGAC 79490  
 |||||

QY 1200 AGTCTAGAAAGTTGTCTACCGCTGTGATCTCAAAAGCTGTGACACACCGTGAGAGAGAA 1259  
 |||||  
 Db 79491 AGTCTAGAAAGTTGTCTACCGCTGTGATCTCAAAAGCTGTGACACACCGTGAGAGAGAA 79550  
 |||||  
 QY 1260 TGAAGAAAGCTGGGCTCTTCAAGGTAAATCTTGTCTTTTCAAAAGCCCTTAATTTTACTGC 1319  
 |||||  
 Db 79551 TGAAGAAAGCTGGGCTCTTCAAGGTAAATCTTGTCTTTTCAAAAGCCCTTAATTTTACTGC 79610  
 |||||  
 QY 1320 ATATATTTTGAATTTCACTGATTAATTTTACAAATTTTCCATPAAGTCAATACACAA 1379  
 |||||  
 Db 79611 ATATATTTTGAATTTCACTGATTAATTTTACAAATTTTCCATPAAGTCAATACACAA 79670  
 |||||  
 QY 1380 TACCTCTCATGCAACACTTGGCTTGTCAATATCATATTTATGAGAGCTGTGCTCT 1439  
 |||||  
 Db 79671 TACCTCTCATGCAACACTTGGCTTGTCAATATCATATTTATGAGAGCTGTGCTCT 79730  
 |||||  
 QY 1440 TAAGGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATTAAGGGATTTGAG 1499  
 |||||  
 Db 79731 TAAGGTAAATGTTTTATATGCACTAAGGCTCTTGGCTTACATATTAAGGGATTTGAG 79790  
 |||||  
 QY 1500 CATGTGATACAGAAAGCTTTTCTCCAGAGTCTCATATGTAAAGAAATTCATTAAGTTGG 1559  
 |||||  
 Db 79791 CATGTGATACAGAAAGCTTTTCTCCAGAGTCTCATATGTAAAGAAATTCATTAAGTTGG 79850  
 |||||  
 QY 1560 CTGAATATAGACTGATCTGTCCATTTCTGTCACTATCATTAAGAGTCAATTAGCTAA 1619  
 |||||  
 Db 79851 CTGAATATAGACTGATCTGTCCATTTCTGTCACTATCATTAAGAGTCAATTAGCTAA 79910  
 |||||  
 QY 1620 GGAAACAAAATCAATCTATGTAAATTAAGAAACAAGCTGTGTTTGTCAATATTAATA 1679  
 |||||  
 Db 79911 GGAAACAAAATCAATCTATGTAAATTAAGAAACAAGCTGTGTTTGTCAATATTAATA 79970  
 |||||  
 QY 1680 TAAGAAAAAGAAACCATGTGAAAGTCAAAATTTGTTTAATCAGTCAATTGAAATCTA 1739  
 |||||  
 Db 79971 TAAGAAAAAGAAACCATGTGAAAGTCAAAATTTGTTTAATCAGTCAATTGAAATCTA 80030  
 |||||  
 QY 1740 TTAATAAGATTTGAATTTTATATGAGAACTATCTGACCAAGTGAACAGTGTGA 1799  
 |||||  
 Db 80031 TTAATAAGATTTGAATTTTATATGAGAACTATCTGACCAAGTGAACAGTGTGA 80090  
 |||||  
 QY 1800 GCTTTTGGCTGTGTGCTCTCAAGTGAAGAGGCTTGTGATTAAGTCTTATATGTGA 1859  
 |||||  
 Db 80091 GCTTTTGGCTGTGTGCTCTCAAGTGAAGAGGCTTGTGATTAAGTCTTATATGTGA 80150  
 |||||  
 QY 1860 CAGGTGCCAAGTTAAGTGTCCCAAGCTTGTCTTAAAGCATCTGATTTTG 1911  
 |||||  
 Db 80151 CAGGTGCCAAGTTAAGTGTCCCAAGCTTGTCTTAAAGCATCTGATTTTG 80202  
 |||||

RESULT 8  
 AC084084/c 71032 bp DNA linear HTG 25-MAR-2001  
 LOCUS  
 DEFINITION Homo sapiens chromosome 8 clone RP11-314C19 map 8, LOW-PASS  
 SEQUENCE SAMPLING.  
 AC084084  
 AC084084.2 GI:13446278  
 KEYWORDS HTG; HTGS\_PHASE0.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 71032)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 8, clone RP11-314C19  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 71032)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abramson,H., Allen,N.,  
 Anderson,S., Bana,N., Bastien,V., Beda,F., Boguslavsky,I.,  
 Boungailier,B., Brown,A., Burkett,G., Campoliano,A., Castelle,A.,  
 Choehei,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,  
 Dearliano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
 Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,

TITLE  
JOURNAL  
COMMENT

Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,  
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRoque, K.,  
Lamaze, R., Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G.,  
Macdonald, P., Marguis, N., McCarthy, M., McEwan, P., McKenna, K.,  
McPheeters, R., Meldrum, J., Meneus, J., Mihova, T., Menga, V.,  
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,  
O'Donnell, P., O'Neill, D., Olivar, T.M., Oliver, J., Peterson, K.,  
Pierre, N., Pisan, C., Pollard, V., Raymond, C., Riback, M., Riley, R.,  
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,  
Sougnuez, C., Spence, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Travers, M., Triglio, J., Vassiliev, H., Viet, R., Vo, A.,  
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
Zimmer, A. and Zody, M.

Direct Submission  
Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 25, 2001 this sequence version replaced gi:10799449.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L11327

Center clone name: 314\_C\_19

-----

\* NOTE: This record contains 85 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 720: contig of 720 bp in length  
\* 721 820: gap of 100 bp  
\* 821 1557: contig of 737 bp in length  
\* 1558 1657: gap of 100 bp  
\* 1658 2404: contig of 747 bp in length  
\* 2405 2504: gap of 100 bp  
\* 2505 3299: contig of 795 bp in length  
\* 3300 3399: gap of 100 bp  
\* 3400 4153: contig of 754 bp in length  
\* 4154 4253: gap of 100 bp  
\* 4254 4978: contig of 725 bp in length  
\* 4979 5078: gap of 100 bp  
\* 5079 5815: contig of 737 bp in length  
\* 5816 5915: gap of 100 bp  
\* 5916 6648: contig of 733 bp in length  
\* 6649 6748: gap of 100 bp  
\* 6749 7520: contig of 772 bp in length  
\* 7521 7620: gap of 100 bp  
\* 7621 8347: contig of 727 bp in length  
\* 8348 8447: gap of 100 bp  
\* 8448 9227: contig of 780 bp in length  
\* 9228 9327: gap of 100 bp  
\* 9328 10052: contig of 725 bp in length  
\* 10053 10152: gap of 100 bp  
\* 10153 10869: contig of 717 bp in length  
\* 10870 10969: gap of 100 bp  
\* 10971 11718: contig of 749 bp in length  
\* 11719 11818: gap of 100 bp  
\* 11819 12557: contig of 735 bp in length  
\* 12558 12657: gap of 100 bp  
\* 12658 13412: contig of 755 bp in length  
\* 13413 13512: gap of 100 bp

13513 14255: contig of 743 bp in length  
\* 14256 14355: gap of 100 bp  
\* 14356 15086: contig of 731 bp in length  
\* 15087 15186: gap of 100 bp  
\* 15187 15901: contig of 725 bp in length  
\* 15902 16001: gap of 100 bp  
\* 16002 16711: contig of 710 bp in length  
\* 16712 16811: gap of 100 bp  
\* 16812 17539: contig of 728 bp in length  
\* 17540 17639: gap of 100 bp  
\* 17640 18351: contig of 712 bp in length  
\* 18352 18451: gap of 100 bp  
\* 18452 19162: contig of 711 bp in length  
\* 19163 19262: gap of 100 bp  
\* 19263 20054: contig of 792 bp in length  
\* 20055 20154: gap of 100 bp  
\* 20155 20891: contig of 737 bp in length  
\* 20892 20991: gap of 100 bp  
\* 20992 21720: contig of 729 bp in length  
\* 21721 21820: gap of 100 bp  
\* 21821 22553: contig of 733 bp in length  
\* 22554 22653: gap of 100 bp  
\* 22654 23420: contig of 767 bp in length  
\* 23421 23520: gap of 100 bp  
\* 23521 24234: contig of 714 bp in length  
\* 24235 24334: gap of 100 bp  
\* 24335 25069: contig of 735 bp in length  
\* 25070 25169: gap of 100 bp  
\* 25170 25915: contig of 746 bp in length  
\* 25916 26015: gap of 100 bp  
\* 26016 26751: contig of 736 bp in length  
\* 26752 26851: gap of 100 bp  
\* 26852 27574: contig of 723 bp in length  
\* 27575 27674: gap of 100 bp  
\* 27675 28437: contig of 763 bp in length  
\* 28438 28537: gap of 100 bp  
\* 28538 29258: contig of 721 bp in length  
\* 29259 29358: gap of 100 bp  
\* 29359 30068: contig of 710 bp in length  
\* 30069 30168: gap of 100 bp  
\* 30169 30918: contig of 750 bp in length  
\* 30919 31018: gap of 100 bp  
\* 31019 31770: contig of 752 bp in length  
\* 31771 31870: gap of 100 bp  
\* 31871 32658: contig of 788 bp in length  
\* 32659 32758: gap of 100 bp  
\* 32759 33501: contig of 743 bp in length  
\* 33502 33601: gap of 100 bp  
\* 33602 34347: contig of 746 bp in length  
\* 34348 34447: gap of 100 bp  
\* 34448 35165: contig of 718 bp in length  
\* 35166 35265: gap of 100 bp  
\* 35266 35998: contig of 733 bp in length  
\* 35999 36098: gap of 100 bp  
\* 36099 36802: contig of 704 bp in length  
\* 36803 37632: contig of 730 bp in length  
\* 37633 37732: gap of 100 bp  
\* 37733 38454: contig of 722 bp in length  
\* 38455 38554: gap of 100 bp  
\* 38555 39275: contig of 721 bp in length  
\* 39276 39375: gap of 100 bp  
\* 39376 40113: contig of 738 bp in length  
\* 40114 40213: gap of 100 bp  
\* 40214 40948: contig of 735 bp in length  
\* 40949 41048: gap of 100 bp  
\* 41049 41792: contig of 744 bp in length  
\* 41793 41892: gap of 100 bp  
\* 41893 42630: contig of 738 bp in length  
\* 42631 42730: gap of 100 bp  
\* 42731 43459: contig of 725 bp in length  
\* 43460 43559: gap of 100 bp  
\* 43560 44277: contig of 718 bp in length



Valas,R., Vera,V., Villaseana,D., Waldton,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Unpublished  
2 (bases 1 to 258662)  
Direct Submission

Worley,K.C.  
Submitted (01-NOV-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 258662)  
Cow Genome Sequencing Consortium.

Direct Submission  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jun 26, 2005 this sequence version replaced gi:58038134.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: FAYB  
Center clone name: CH240-2L12

----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 214934 bases at least Q40  
Consensus quality: 217760 bases at least Q30  
Consensus quality: 220382 bases at least Q20  
Estimated insert size: 219583; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 30 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
3528 3527: contig of 3527 bp in length  
3578 3577: gap of 50 bp  
3578 5414: contig of 1837 bp in length  
5415 5464: gap of 50 bp  
5465 6778: contig of 1314 bp in length  
6779 8809: gap of 2031 bp  
8810 9920: contig of 1111 bp in length  
9921 10020: gap of unknown length  
10021 12558: contig of 2538 bp in length  
12559 12608: gap of 50 bp  
12609 17883: contig of 5275 bp in length  
17884 17983: gap of unknown length  
17984 30723: contig of 12740 bp in length

30724 30773: gap of 50 bp  
30774 46311: contig of 1538 bp in length  
46312 46361: gap of 50 bp  
46362 54121: contig of 7760 bp in length  
54122 54171: gap of 50 bp  
54172 69599: contig of 15428 bp in length  
69600 69649: gap of 50 bp  
69650 76463: contig of 6843 bp in length  
76463 76512: gap of 50 bp  
76513 78033: contig of 1521 bp in length  
78034 78133: gap of unknown length  
78134 81273: contig of 3140 bp in length  
81274 81323: gap of 50 bp  
81324 92153: contig of 10630 bp in length  
92154 92254: gap of unknown length  
92254 93265: contig of 1012 bp in length  
93266 94230: gap of 965 bp  
94231 115783: contig of 21553 bp in length  
115784 115833: gap of 50 bp  
115834 118884: contig of 3051 bp in length  
118885 118934: gap of 50 bp  
118935 151000: contig of 32066 bp in length  
151001 151050: gap of 50 bp  
151051 152191: contig of 1141 bp in length  
152192 182104: gap of 29913 bp  
182104 189192: contig of 7088 bp in length  
189193 189242: gap of 50 bp  
189243 192962: contig of 3720 bp in length  
192963 193012: gap of 50 bp  
193013 209262: contig of 16550 bp in length  
209263 209312: gap of 50 bp  
209313 231496: contig of 22184 bp in length  
231497 231767: gap of 271 bp  
231768 233967: contig of 2200 bp in length  
233968 234067: gap of unknown length  
234068 235110: contig of 1043 bp in length  
235111 235210: gap of unknown length  
235211 236716: contig of 1506 bp in length  
236717 236816: gap of unknown length  
236817 238176: contig of 1360 bp in length  
238177 238276: gap of unknown length  
238277 240200: contig of 1924 bp in length  
240201 240300: gap of unknown length  
240301 242112: contig of 1812 bp in length  
242113 242212: gap of unknown length  
242213 242213 258662: contig of 16450 bp in length.

FEATURES  
source  
1..258662  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-2L12"  
3528..3577  
/estimated\_length=50  
5415..5464  
/estimated\_length=50  
6779..8809  
/estimated\_length=2031  
9921..10020  
/estimated\_length=unknown  
12559..12608  
/estimated\_length=50  
17884..17983  
/estimated\_length=unknown  
30724..30773  
/estimated\_length=50  
46312..46361  
/estimated\_length=50  
54122..54171  
/estimated\_length=50  
69600..69649  
/estimated\_length=50  
76463..76512

Query Match 20.3%; Score 407; DB 14; Length 258662;  
 Best Local Similarity 66.9%; Pred. No. 2.9e-79;  
 Matches 814; Conservative 0; Mismatches 240; Indels 163; Gaps 10;

QY 935 CTCGAATACAGAGTGTCTAGAGAGCTGTATTTGGGCACTTAACATATTCCTC-----A 987  
 DB 58490 CTAAATGTGAATGTTCTGGAGAGCACTATTTGGGCACTTAATGCTTCTCCAAAAGT 58431

QY 988 CTACTTAGGGCAGACACT-GAAATTAACCACTAAGTTGCAATGCTCATGTAGT 1046  
 DB 58430 TTGCTTAGGGCAGACAGCTGAATTAACACTCTATGCTGCATATCCATGGGGTTGG 58371

QY 1047 TCTCAGGAGCTGAGGCTCAGAGAGTGAACCTGATATGATGTCAGGCTTCTTC 1106  
 DB 58370 TCTCAGGAGCAGACAGCTTCAGAGTGAACCAATCTCTGTTGGCAGGCTTCTCTCC 58311

QY 1107 CTTCAGAGTCACTGTTGTTCTGCTGACTCTCCATAGAGACATGAGT--CCTGAATCC 1164  
 DB 58310 CTTCAGAGTCACTGTTGTTCTGCTGACTCTCCATAGAGAGAGAGTCCGATCCCT 58251

QY 1165 TCAGACCACTATCTGAGTATGTAAGTCTCTGACAGTCTTGAAGTTGTTACCGCTGG 1224  
 DB 58250 CAAACCATCACTCGAGTATGTAAGTCTGACAGTCCAGAGAAAGCATTCAGTCTTCG 58191

QY 1225 ATCTCAGAGCGGTGACACACCGTGAGAGAGAAATGAAGAGCTGGCTTCAGAGTAA 1284  
 DB 58190 AGCTGAGAGACCTGTATGATTAATTAAGAGAGTGAAGAGAAAGCTGAGCTTTCAGATTA 58131

QY 1285 ATCTGCTTTTTCACAAGCCCCCTTAATTTTACATGCAATTAATTTTGAATTCAGTATA 1344  
 DB 58130 ACCTTGTTTT--ACAAGTCCCTCATTTTATGCAATTAATTTTGAATTCAGTATA 58073

QY 1345 TTTCTCAATTTTCCATAAGTATGATTCACACAAATACCTCTCATGCAACACTGGCTT 1404  
 DB 58072 TTTGTACAATTTTCTATAAGTATGATTCACACAAATCTCTCCATGAGACACTGGCTT 58013

QY 1405 TGTATATACATATCTAT-ATGAGAGCTGCTCTTAAAGCGTAATGTTTATATGAC 1463  
 DB 58012 TGTATATACATATCTAT-ATGAGAGCTGCTCTTAAAGCGTAATGTTTATATGAC 57953

QY 1464 TAAGGCTCTGCTCATATTAAGAGGAGTATGAGCAATGTATACAGAGCTTTTCT 1523  
 DB 57952 TAAGGCTCTGCTCATATTAAGAGGAGTATGAGCAATGTATACAGAGCTTTTCT 57895

QY 1524 CCACAGGCTCATATGTAAGAGTATGATTTGCTGAATGAGCTGATCTGTCAT 1583  
 DB 57894 CTA-----CACATATCAAGATTCATGGGATT-GCTGAGATGAGCTGATTTATCTATT 57843

QY 1584 TCTCTGCTCATATTAAGAGAGTATGAGTAAAGAAACAAACTACATCTATGTA 1643  
 DB 57842 TCTCTGCTCATATTAAGAGAGTATGAGTAAAGAAACAAACTACATCTATGTA 57783

QY 1644 ATTAGAAGAACAGAGTGTGTTTCTCAATTAATAAATAAAGAAACCAATGTAAG 1703  
 DB 57782 GTTAAATGAACAGAGTGTGTTTCTCAATTAATAAATAAAGAAACCAATGTAAG 57724

QY 1704 TCAAAATATTTGTTTATACAGTCAATGAGATCTAT----- 1741  
 DB 57723 AGACAGATTTTCTTATACAGTCAATGAGATCTAT----- 57664

QY 1742 ----- 1741  
 DB 57663 ATATATATATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 57604

QY 1742 -----AAA 1744  
 DB 57603 ATATATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 57544

QY 1745 AAGTATTTGAATCTTTATGATGAGAACTATCTTGAAGTGAAGAGTGTGAGCTTT 1804  
 DB 57543 TAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 57484

QY 1805 TTGGCTGTGTGCTCCCTACGTAGAAGAGAGCTTTGTCATATAAGTCTTATATGATACAGT 1864  
 DB 57483 CTGACTGCTGTATATTTACATAGATATGATGCTTTGGCATAAATCTCATCTACAGAGAC 57424

QY 1865 GCCAAGTTAAGTGGCCCAAGCTTCTTAAAGCATGAGATTTGTTTGAACCTTTTA 1924  
 DB 57423 ACTGAATAATAGTCCCTGAGCTGCTTTTAAAGCACTGAGATTAATTAATCTTCA 57364

QY 1925 GTGAATGAAGGAATAAACAATCCCTCTGGAGAACTTCTCCATCCCTTGGGAAG 1984  
 DB 57363 GTGAATGAAGGAATAAACAATCCCTCTGGAGAAACCTTCCATCCCTTGGGAAG 57304

QY 1985 TCATCTGCCAGAAATTC 2001  
 DB 57303 TCATCTGCCAGAAATTC 57287

RESULT 10  
 AC134591  
 LOCUS  
 DEFINITION Mus musculus BAC clone RP23-376L4 from chromosome 19, complete sequence.  
 ACCESSION AC134591  
 VERSION AC134591.4 GI:38348187  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Vanbrunt, A., Van Brunt, A., Cotton, M., Meyer, R. and Haakenson, W.  
 TITLE The sequence of Mus musculus BAC clone RP23-376L4  
 JOURNAL Unpublished (2001)  
 REFERENCE 2 (bases 1 to 222814)  
 AUTHORS Wilson, R.  
 TITLE Sequencing of Mus musculus  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 222814)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submision  
 JOURNAL Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submision  
 JOURNAL Submitted (28-DEC-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
 AUTHORS Wilson, R.K.  
 TITLE Direct Submision  
 JOURNAL Submitted (15-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
 AUTHORS Wilson, R.  
 TITLE Direct Submision  
 JOURNAL Submitted (01-JAN-2004) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Nov 15, 2003 this sequence version replaced gi:27413228.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUSC  
 Web site: http://genome.wustl.edu  
 Contact: submissions@wustl.edu  
 ----- Summary Statistics  
 Center project name: M\_BA0376L04

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minko Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.

FEATURES  
Location/Qualifiers

```
1..222814
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="19"
/map="19"
/clone="RP23-376L4"
/clone_1lb="RPCI-23"
1..55
/repeat_region
/rpt_family="Alu"
147..265
/rpt_family="L1"
404..543
/rpt_family="MaLR"
528..841
/rpt_family="MaLR"
1425..1598
/rpt_family="B2"
1635..1800
/rpt_family="L1"
2314..2430
/rpt_family="Alu"
2986..3130
/rpt_family="B4"
4371..4534
/rpt_family="MIR"
5061..5126
/rpt_family="MER2_type"
6039..6273
/rpt_family="B4"
6287..6631
/rpt_family="MaLR"
6642..6790
/rpt_family="B4"
7481..7545
/rpt_family="ERV1"
8637..8698
/rpt_family="Alu"
8773..9062
/rpt_family="B4"
9142..9552
/rpt_family="MaLR"
9553..9600
/rpt_family="ID"
9601..9615
/rpt_family="B4"
10198..10331
/rpt_family="Alu"
```

```
repeat_region 10429..10514
/rpt_family="ERVK"
repeat_region 10470..10902
/rpt_family="ERVK"
repeat_region 12643..12771
/rpt_family="B4"
repeat_region 16204..16251
/rpt_family="ERV1"
repeat_region 20296..20440
/rpt_family="Alu"
repeat_region 20738..21229
/rpt_family="ERV1"
repeat_region 26642..27133
/rpt_family="ERV1"
repeat_region 27205..27409
/rpt_family="B2"
repeat_region 27445..27588
/rpt_family="B4"
repeat_region 27764..27813
/rpt_family="ERV1"
repeat_region 27917..28095
/rpt_family="ERVK"
repeat_region 28021..28327
/rpt_family="MaLR"
repeat_region 28540..28660
/rpt_family="MaLR"
repeat_region 28897..29037
/rpt_family="Alu"
repeat_region 29352..29504
/rpt_family="Alu"
repeat_region 29543..29676
/rpt_family="Alu"
repeat_region 31118..31199
/rpt_family="L2"
31652
/note="Unresolved bases."
32320..32524
/rpt_family="B2"
32542..32598
/rpt_family="Alu"
34883..35188
/rpt_family="MaLR"
35500..35695
/rpt_family="MaLR"
35787..36037
/rpt_family="MaLR"
36446..36625
/rpt_family="MaLR"
37694..38232
/rpt_family="L1"
39171..39242
/rpt_family="Alu"
39465..39647
/rpt_family="MIR"
39648..39764
/rpt_family="Alu"
39765..39827
/rpt_family="MIR"
39942..40150
/rpt_family="B2"
40468..40550
/rpt_family="L2"
40619..41063
/rpt_family="MaLR"
41084..41408
/rpt_family="MaLR"
41409..41757
/rpt_family="MaLR"
41948..42018
/rpt_family="L2"
42080..42185
/rpt_family="MIR"
43881..44239
```







Query Match	4.9%	Score 97.6	DB 6	Length 300
Best Local Similarity	56.7%	Pred. No. 9.7e-11		
Matches	168	Conservative 18	Mismatches 62	Indels 4
				Gaps 4
OY	1753	GAATTCCTTATGATGAGAACTATCTTGA	CTCAAGTGAAGCAAGTGTGACTTTTGGCCTG	1812
Db	1	GAATTCCTGKWKATTAKAACTATCTTG	CTCAAAATTSACTGGTAGCTAACCTGGCCTG	60
OY	1813	TGCTCCCTACG-TAGAAAGAGGCTTTGT	CATAAG-TCTTATATGTACAGGTGCCAAG	1870
Db	61	TGCTCCCTGCTTTAAGAGAGGCTTTGT	CTATATGATCACTMTGTGACTACGTGCCCTAG	120
OY	1871	TT-ANGTCCCCAAGCTTGCTCTCTTA	AAACAATCTGATTTTTGTTTATGACTTTTAG	1929
Db	121	TTGTATGTCCCTGCTTGTCTGTCTT	CTTCTGAGCTTATTTGGGTATATACATC	180
OY	1930	CTGAAGGAAATTAACAAATCCCTCT	GGAGAGACTTCTCTCCATCTTGTTGAATCA	1989
Db	181	YTSAAAGGTCT-TTCTCCTCGG	YGGAGAAATTTCTCTCTCTCTCGAGAACTCTT	239
OY	1990	CTGCCAGAAATTC	2001	
Db	240	CTSCCGAAATTC	251	
RESULT 13				
LOCUS	166494	166494	7218 bp	DNA linear PAT 28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.			
ACCESSION	166494			
VERSION	166494.1	GI:2724471		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 7218)			
AUTHORS	Donner, F., Schefflinger, F. and Falkner, F. Gunter.			
TITLE	Recombinant fowlpox virus			
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;			
FEATURES	Location/Qualifiers			
source	1..7218			
	/organism="unknown"			
ORIGIN	/mol_type="unassigned DNA"			
Query Match	3.0%	Score 60.8	DB 6	Length 7218
Best Local Similarity	2.9%	Pred. No. 0.0085		
Matches	11	Conservative 226	Mismatches 143	Indels 0
				Gaps 0
OY	408	CAGAAAAAAGAACATAGGAGCTAC	TGCGGGCTTTTANCTTAAGCCTTGATCTTC	467
Db	1046	CAGGTGAGGAGCTTGCGATTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1105
OY	468	TTTTACAGCTTACTACCTGCTTG	ACTGAGCCATATGCCCTGTAAAGCTTC	527
Db	1106	TTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1165
OY	528	GATTCTTACTAGCTGGGCTTCT	TATATAGCTCTTCTCCATTTCTGTGCTCA	587
Db	1166	TTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1225
OY	588	TGATCTTCTCTTTTCTCACCTC	TGGAGCTGGGCTGTTGTATGAGACGCTTA	647
Db	1226	TTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1285
OY	648	TTGCTTTGGTTTTTCTCTGG	GGACAATGTCTCAGATTAATCTAGACCA	707
Db	1286	TTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1345
OY	708	AGCACTGGGCGAGGCTTCT	CTCTCCAATCGACATGTTCCAGGCTCT	767
Db	1346	TTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1405

QY	768	AGTTAGGTCAAGATTCCTT	787
Db	1406	YYYYYYYYYYYYYYYYYY	1425
RESULT 14			
AX655393	LOCUS	AX655393	2000 bp DNA
AX655393	DEFINITION	Sequence 5263 from Patent WO03000898.	linear
AX655393	ACCESSION	AX655393	
AX655393.1	VERSION	AX655393.1	
GI:29158207	KEYWORDS		
ORGANISM		Oryza sativa	
SOURCE		Oryza sativa	
REFERENCE		Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M., Kaegebein, F., Quan, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G. Plant genes involved in defense against pathogens Patent: WO 03000898-A 5263 03-JAN-2003; Syngenta Participations AG (CH)	
FEATURES		Location/Qualifiers	
source		1..2000	
ORIGIN		/organism="Oryza sativa"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:4530"	
Query Match	2.9%	Score 58.8; DB 6; Length 2000;	
Best Local Similarity	11.8%	Pred. No. 0.027;	
Matches	54; Conservative	215; Mismatches 187; Indels 2; Gaps 1;	
QY	1361	ATAAGTCATCTACACAAATACCCCTCATGCAACACTTGGCTTGTCTAATACATATCTA	1420
Db	508	MYRMRKRYMYSARAYTRKRCARKKYSYSAARAKACWRGKGYWAGMMRKRYKMYM	567
QY	1421	TTATGAGAGCTGCTCTTCTTAAGGTTAAATGTTTATATGACCTAAGGCTTGGCTTAC	1480
Db	568	MMWYKRYKSKSWYCKMSYVASCMSKARAKGAKCKRSKMSAWSRSSRCKRCKASKR	627
QY	1481	ATATATAAGGGGTATTAGACATGTGATPACGAAGTCTTTTCTCACAGGCTCATATGT	1540
Db	628	SSARRYAMGMSGSSMSRKSTCTCYMRKMSKSTCTMYTMSKTYAKGSYMRRY	687
QY	1541	AAAGAATTCATTAGATTGGCTGAATAGACTGATCTGTCCATTTCTGTGCTCACTTATCA	1600
Db	688	RAMCYMMRWYRYRYSYMTYMAVYTSSTRAMTGMKYSGRWYTSWYKCKSCSKYRSW	747
QY	1601	TAAAGAACTATTAGCTACGACAAACAAACATCAATCTATGTAATTGAAAGAACAACTG	1660
Db	748	YYSMMWMAKTKMWRRAATRRMMWYYSKMKWTYCTMGWYMMWTYMKRMRYM--YK	805
QY	1661	GTTTTGCTCAATATAAATAAGAAAGAAACCATGTGAAGTCAATATATTTGTTAA	1720
Db	806	CTKTYWYSATYWTGTBAAMMMATIKRBMGHTGAKTRGRARARYYMKATWCTKRMWT	865
QY	1721	TCAAGTCATTGAGAAATCTATTTAAAGATATTGTAATCTTTATGAGAGACTATCTTGA	1780
Db	866	KGKAKXAMWTMAKAMKRYWMSWRAMVYKYTRTRTYKTCWMAKARFSGWAWYMMWKS	925
QY	1781	CTCAAGTGAGACAGTGTGAGCTTTTGGCTGTGCTGC	1818
Db	926	AKMWMWKGGRMGWTKYWTCTTWMKACGRATKIMCC	963
RESULT 15			
LOCUS	166494	7218 bp DNA	linear
DEFINITION	Sequence 14 from patent US 5670367.		
ACCESSION	166494		

KEYWORDS	166494.1	GI:2724471
VERSION	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 7218)	
AUTHORS	Donner, F., Schefflinger, F. and Falkner, F. Gunter.	
TITLE	Recombinant fowlpox virus	
JOURNAL	Patent: US 5670367-A 19 23-SEP-1997;	
FEATURES	Location/Qualifiers	
source	1..7218	
ORIGIN	/organism="unknown" /mol_type="unassigned DNA"	
Query Match	2.9%; Score 58.4; DB 6; Length 7218;	
Best Local	Similarity 1.4%; Pred. No. 0.029;	
Matches	5; Conservative 216; Mismatches 127; Indels 0; Gaps 0;	
QY	8 TTTGTATCAGAAAAAGAAAAAGACAGCTCAAAATATATTAACGTCATATAGTTAGTACAGAC 67	
DB	1441 TTGGTACACRR 1382	
QY	68 CATGATTGACAGTACAGAACTCAATCAACCAACGTAAGTCAAAGAGAAATATAT 127	
DB	1361 RRR 1322	
QY	128 TGGCTCATGTATCCTTCTCACAGAGAGGACAGATGAGAGGGCTTTGGAGAACAGAA 187	
DB	1321 RRR 1262	
QY	188 TTGTTCTCAAAATTTAGAGAAATCTAGATTTAGTCAGAGATGGCTCCTTCTGCTG 247	
DB	1261 RRR 1202	
QY	248 AGGTGCTGTCGTCAGTGTAGAGTCTTATGGAGAGAGAGTGCATGTAGATGAAGGT 307	
DB	1201 RRR 1142	
QY	308 AGGCTAAGCAAAACAGGCAAGGCGCCACTATATCATCTAAATATG 355	
DB	1141 RRR 1094	
RESULT 16		
AL137881/C	143324 bp	Linear PRI 18-MAY-2005
LOCUS	Human DNA sequence from clone Rp11-40A8 on chromosome 13. Contains a novel gene, the 3' end of the GUCY1B2 gene for guanylate cyclase 1 soluble beta 2 and a Cpg island, complete sequence.	
DEFINITION	AL137881	
ACCESSION	AL137881.12	GI:8894206
VERSION	HTG; Cpg island; guanylate cyclase; GUCY1B2.	
KEYWORDS	Homo sapiens (human)	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 143324)	
AUTHORS	Pelan, S.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk	
COMMENT	Clone request: clonerequest@sanger.ac.uk On Jul 1, 2000 this sequence version replaced gi:8546598. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information from the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep/ This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at	

```

FEATURES
      http://www.sanger.ac.uk/HGP/Chr13
      RP11-40A8 is from the library RGC1-11.1 constructed by the group of
      Pieter de Jong. For further details see
      http://www.chori.org/bacpac/home.htm
      VECTOR: pBACE3.6
      ----- Genome Center
      Center: Wellcome Trust Sanger Institute
      Center code: SC
      Web site: http://www.sanger.ac.uk
      Contact: vega@sanger.ac.uk
      -----
      This sequence was finished as follows unless otherwise noted: all
      regions were either double-stranded or sequenced with an alternate
      chemistry or covered by high quality data (i.e., phred quality >=
      30); an attempt was made to resolve all sequencing problems, such
      as compressions and repeats; all regions were covered by at least
      one subclone; and the assembly was confirmed by restriction digest,
      except on the rare occasion of the clone being a YAC.
      Location/Qualifiers

```

```
source    .._l43j3z4  
          /organism="Homo sapiens"  
          /mol_type="genomic DNA"  
          /db_xref="taxon:9606"  
          chromosome="13"  
          clone="RP11-40A8"  
          /clone_lib="RPCI-11.1"  
  
misc_feature  
      1  
         /note="Clone_left_end: RP11-40A8"  
  
misc_feature  
      14112  
         /note="Clone_right_end: RP11-233H19"  
  
misc_feature  
      19526..19591  
         /note="Single clone region. Region contains reads from  
clone port only (Oligos 1&2). Assembly is consistent with  
restriction digest."  
  
gene     complement(join(25997..26264,33228..33554,54102..54331))  
          /locus_tag="RP11-40A8.3-.001"  
          /locus_tag="RP11-40A8.3-.001"  
          complement(join(25997..26264,33228..33554,54102..54331))  
          /locus_tag="RP11-40A8.3-.001"  
          /locus_tag="RP11-40A8.3-.001"  
  
mrna     /note="match: ESTs: BF574759.1"  
          join(SJ310..53759,71026..71097,73094..73201,74302..74378,  
              78504..78618,86940..87013,89046..89151,91606..91687,  
              93082..93124,97524..97604,99977..100383)  
          /locus_tag="RP11-40A8.1-.001"  
          join(SJ310..53759,71026..71097,73094..73201,74302..74378,  
              78504..78618,86940..87013,89046..89151,91606..91687,  
              93082..93124,97524..97604,99977..100383)  
  
mrna     /note="match: ESTs: AL520828.ALU539901.ALU568105.AUU19164  
BG25097.BG259492.BG261287.BG281634.BG393104.BG724123  
B1460913.B1754776.BM906357.BM918887.BM999529.BQ016444  
BQ216110  
match: CDNAs: AK021774.BC001397.BC005088.BC007332  
BC010174"  
          join(SJ310..53759,56614..56748,71024..71085)  
          /locus_tag="RP11-40A8.1-.002"  
          join(SJ310..53759,56614..56748,71024..71085)  
  
gene     /locus_tag="RP11-40A8.1-.002"  
          /locus_tag="RP11-40A8.1-.002"  
          /note="match: ESTs: BM847956.1"  
          join(SJ310..53759,56614..56748,71024..71097,73094..73201,  
              74302..74378,86940..87002)  
          /locus_tag="RP11-40A8.1-.005"  
          /locus_tag="RP11-40A8.1-.005"  
mrna     /note="match: ESTs: B1753018.1"  
          join(SJ396..53759,71026..71097,73094..73201,74302.  
              74378,86940..87013,89046..89151,91606..91687,  
              93082..93124,97524..97604,99977..>100090)  
          /standard_name="OTTHUMP0000001436"  
          /note="match: proteins: Q9D014.Q9NAHF"  
          /codon_start=1  
          /protein_id="CAII3913.1"
```

/db\_xref="GI:55957387"  
/translation="MAAGVDCGDGVGAROHFVLVSEYLKQASKKRNGLMFVKLVNPG  
SGEGAIYLPNNCLQOLFVKKVFKKSHFVFNVSOGGLHFRATVPDLPFLIHLI  
KADKGGKIPDQVNVNVPNCILKLLGLKEJLLHVEKGNEDINKKYKYSK  
KYLITMKKQNOIVAAIKTNVNVSVRSVSTFSPSDASTDKEEDYIRYAHGLSD  
YTPKRLSDYKYLKLEPRASLPNPSSKIKLSDPEVKAKEDYTFPNKDLKTEKKN  
SKMTPAQALAKVDSGKMSIDTFPGYKNNKKIKGV"  
69479  
/misc\_feature  
/locus\_tag="RP11-40A8.1-002"  
/note="Clone left end: RP11-547C18"  
join(71981, -72127, 73094, -73201, 74302, -74378, 78504, -78618, 86940, -87013, 89046, -89151, 93080, -93124, 97524, -97604)  
/locus\_tag="RP11-40A8.1-004"  
join(71981, -72127, 73094, -73201, 74302, -74378, 78504, -78618, 86940, -87013, 89046, -89151, 93080, -93124, 97524, -97604)  
/locus\_tag="RP11-40A8.1-004"  
/note="match: ESTs: BM55965.1"  
join(73733, -73774, 74302, -74378, 86940, -87013, 89046, -89133)  
/locus\_tag="RP11-40A8.1-003"  
join(73733, -73774, 74302, -74378, 86940, -87013, 89046, -89133)  
/locus\_tag="RP11-40A8.1-003"  
/note="match: ESTs: BE972426.1"  
join(93101, -93124, 93690, -93337, 97524, -97604, 99977, -100121)  
/locus\_tag="RP11-40A8.1-006"  
join(93101, -93124, 93690, -93337, 97524, -97604, 99977, -100121)  
/locus\_tag="RP11-40A8.1-006"  
/note="match: ESTs: AM439381.1"  
100360, -100365  
/locus\_tag="RP11-40A8.1-001"  
100383  
/locus\_tag="RP11-40A8.1-001"  
join(complement(AL160157, -17:65346, -66401),  
complement(AL160157, -17:63216, -63289),  
complement(AL160157, -17:48615, -48709),  
complement(AL160157, -17:34398, -34554),  
complement(AL160157, -17:29989, -30088),  
complement(AL160157, -17:28603, -28799),  
complement(AL160157, -17:27097, -27208),  
complement(AL160157, -17:25729, -25832),  
complement(AL160157, -17:24495, -24711),  
complement(AL160157, -17:20793, -21113),  
complement(AL160157, -17:16576, -16722),  
complement(AL160157, -17:10468, -10622),  
complement(AL160157, -17:7385, -7536),  
complement(AL160157, -17:7095, -7304),  
complement(AL160157, -17:6648, -6971),  
complement(138130, -138430)  
/locus\_tag="RP11-40A8.2-001"  
join(complement(AL160157, -17:66346, -66401),  
complement(AL160157, -17:63216, -63289),  
complement(AL160157, -17:48615, -48709),  
complement(AL160157, -17:34398, -34554),  
complement(AL160157, -17:29989, -30088),  
complement(AL160157, -17:28603, -28799),  
complement(AL160157, -17:27097, -27208),  
complement(AL160157, -17:25729, -25832),  
complement(AL160157, -17:24495, -24711),  
complement(AL160157, -17:20793, -21113),  
complement(AL160157, -17:16576, -16722),  
complement(AL160157, -17:10468, -10622),  
complement(AL160157, -17:7385, -7536),  
complement(AL160157, -17:7095, -7304),  
complement(AL160157, -17:6648, -6971),  
complement(138130, -138430)  
/locus\_tag="RP11-40A8.2-001"  
/product="guanylate cyclase 1, soluble, beta 2"  
/note="Match: ESTs: A1247180 A1792818 A1822009 AN300728  
BG181663 BG680126.1

ORIGIN	misc_feature	match: cDNAs: AF030499"
Query Match	2.8%; Score 56.4; DB 8; Length 143324;	
Best Local Similarity	62.7%; Pred. No. 0.055;	
Matches 104; Conservative	0; Mismatches 61; Indels 1; Gaps 1;	
polyA_site	match: cDNAs: AF030499"	
polyA_signal	complement (138130) /gene="GUCY1B2" /locus_tag="RP11-40A8.2-001" complement (138150, 138155) /gene="GUCY1B2" /locus_tag="RP11-40A8.2-001" 143324 /note="Clone_right_end: RP11-40A8"	
Db	18 AAAAAAAAAAGAGTGTGAAAAATATATATTAACCTGCATATAGTTCCAGACCATGATTCG 77	
Qy	33346 AAAAAAGTGTGAGCAGCTGCATTAAGTATCTGTATTAAGCAGGGTATATATGATTGC 3328	
Db	78 AAGTACAGAAACTCAAT-TCAAAACCAACCTAAGTCACAAAAGAAATATATTTGGCTCATG 136	
Qy	33286 AAGTACAGAAACCAATCAGAAATCACTGAGAGTCATGAGGAACACTATTGTGCTCAG 3322	
Db	137 TAACCTTCTCAGAGAGGGCAGAGATGGAAGGGGCTTTGGGAAACA 132	
Qy	33226 TAACTTAACTTGAATGGCAGAGGATGTATCTGTTTGGGAGGAA 33181	
RESULT 17		
LOCUS	AC021786	
DEFINITION	Homo sapiens clone RP11-21H8, WORKING DRAFT SEQUENCE, 8 unordered pieces.	
ACCESSION	AC021786	
VERSION	AC021786.2 GI:7341964	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 163554) Britten, B., Linton, L., Nussbaum, C. and Lander, E. Homo sapiens, clone RP11-21H8 Unpublished	
AUTHORS	2 (bases 1 to 163554) Britten, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barina, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collimore, A., Cooke, P., Dattelaro, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lander, S. T., Lehoczek, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olyva, J. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Sotjanovic, N., Subramanian, A., Talamas, J., Testfay, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.	
REFERENCE	Direct Submission Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Mar 30, 2000 this sequence version replaced gi:6721341. All repeats were identified using RepeatMasker: Smit, A. F. A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/XM/RepeatMasker.html	
TITLE	Genome Center	
JOURNAL	Center: Whitehead Institute/ MIT Center for Genome Research	
COMMENT	Center code: WIBR	

misc_feature	73305..109834	/note="assembly_fragment"
gap	109835..109934	/estimated_length=100
misc_feature	109935..163554	/note="assembly_fragment"
	clone_end:T7	vector_side:right"
Query Match	2.8%; Score 56.4; DB 14; Length 163554;	
Best Local Similarity	62.7%; Pred. No. 0.054;	
Matches 104; Conservative	0; Mismatches 61; Indels 1; Gaps 1;	
Db	47966	TAACCTTAACCTTGAAGCAGCTCAGATGAACTCTGTATTATGACAGGATATATGATTATGTC 480255
Qy	78	AAGTACAGAACTCAAT-TCAACCAAGTAAAGTCAAAAGGAAATATATATGCGTCATG 136
Db	48026	AAGTACAGAACTCAATGAACTCACTGAGTATAGAGAACCTTATTTGGCTCAGG 480855
Qy	137	TAACTTTCTCAGAGAGGCGAGATGAGAGGGCTTTGGGACAA 182
Db	48086	TAACTTAACCTTGAATGGCAGGAGATGATCTGTTGGAGGAA 48131
RESULT 18		
AC008180	193035 bp	DNA linear PRI 29-MAY-2002
LOCUS	Homo sapiens 3 BAC Rpl1-29225 (Roswell Park Cancer Institute Human	
DEFINITION	BAC library) complete sequence.	
VERSION	AC008180	
KEYWORDS	AC008180.15 GI:9558559	
SOURCE	HTG.	
ORGANISM	Homo sapiens (human)	
REFERENCE		
AUTHORS	1 (bases 1 to 193035)	
	Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,	
	Bodoto,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunn,C.,	
	Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,	
	David,R., Delgado,O., Deshaun,D., Ding,Y., Domak-Rashid,N.,	
	Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferriguto,D.,	
	Forcun-Tansey,J., Frantz,P., Ganesher,R., Garcia,D.K., Gorrell,J.H.,	
	Gorrell,L.L., Guevara,W., Harris,K., He,X., Hernandez,J.,	
	Hodgson,A., Hognes,M., Holmway,C., Hosak,H., Jackson,L.B.,	
	Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondjewski,N., Kong,Y.,	
	Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O.,	
	Lozadó,R., J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P.,	
	Mejza,M., Moore,S., Moorish,T., Morgan,M., Morris,S., Nash,S.,	
	Nelson,A., Nguyen,R., Nguyen,S., Oswal,G., Parish,B.,	
	Patton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D.,	
	Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H.,	
	Simon,M., Sparks,A., Stamps,A., Sugang,R., Taber,P., Taylor,T.,	
	Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,	
	Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,	
	Wrenstock,G., Yu,W., Zhou,X., Naylor,S.L., Nelson,D. and Gibbs,R.	
TITLE	Direct Submission	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 193035)	
AUTHORS	Worley,K.C.	
TITLE	Direct Submission	
JOURNAL	Submitted (29-JUL-1999) Human Genome Sequencing Center, Department	
	of Molecular and Human Genetics, Baylor College of Medicine, One	
	Baylor Plaza, Houston, TX 77030, USA	
REFERENCE	3 (bases 1 to 193035)	
AUTHORS	Worley,K.C.	
TITLE	Direct Submission	
JOURNAL	Submitted (28-JUL-2000) Human Genome Sequencing Center, Department	
	of Molecular and Human Genetics, Baylor College of Medicine, One	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 193035)  
Morley, K.C.  
Direct Submission  
Submitted (22-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 193035)  
Morley, K.C.  
Direct Submission  
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 28, 2000 this sequence version replaced gi:9438809.  
COMMENT  
JOURNAL  
gc-help@bcm.tmc.edu  
http://www.hgsc.bcm.tmc.edu/ or email

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL:  
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.  
location/Qualifiers  
source  
1. .193035  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-292L5"  
1. .2005  
/note="overlaps bases 88684. .90688 of clone AC119732"  
/function="Clone overlap"  
1. .1169  
/rpt\_family="L1"  
repeat\_region  
complement(1172. .1396)  
/rpt\_family="AluJo"  
repeat\_region  
complement(1398. .1800)  
/rpt\_family="L1"  
repeat\_region  
complement(1811. .2094)  
/rpt\_family="L1M2"  
2236. .2775  
/rpt\_family="L1MA4A"  
repeat\_region  
complement(2820. .3029)  
misc\_feature  
1. .2005  
/note="overlaps bases 88684. .90688 of clone AC119732"  
/function="Clone overlap"  
1. .1169  
/rpt\_family="L1"  
repeat\_region  
complement(1172. .1396)  
/rpt\_family="AluJo"  
repeat\_region  
complement(1398. .1800)  
/rpt\_family="L1"  
repeat\_region  
complement(1811. .2094)  
/rpt\_family="L1M2"  
2236. .2775  
/rpt\_family="L1MA4A"  
repeat\_region  
complement(2820. .3029)

repeat\_region  
/rpt\_family="L1MC4"  
3030. .3081  
/rpt\_family="TRGA)n"  
complement(3082. .3242)  
repeat\_region  
/rpt\_family="L1MC4"  
complement(3496. .3890)  
/rpt\_family="LTR16A"  
complement(3953. .4079)  
repeat\_region  
/rpt\_family="MER3"  
complement(5947. .6007)  
repeat\_region  
/rpt\_family="MER94"  
6130. .6505  
/rpt\_family="MLT1B"  
7376. .7410  
/rpt\_family="(A)n"  
7414. .7467  
/rpt\_family="(CA)n"  
complement(9799. .10540)  
/rpt\_family="L1MB6"  
complement(10543. .10958)  
/rpt\_family="L1M4"  
11299. .11329  
/rpt\_family="(TG)n"  
12452. .12472  
/rpt\_family="(T)n"  
complement(13598. .13880)  
/rpt\_family="MIR"  
13925. .13951  
/rpt\_family="AT\_rich"  
13989. .14014  
/rpt\_family="AT\_rich"  
14926. .14974  
/rpt\_family="AT\_rich"  
15078. .15108  
/rpt\_family="(CA)n"  
complement(15519. .15583)  
/rpt\_family="MIR"  
complement(15702. .15859)  
/rpt\_family="MIR"  
complement(16543. .16826)  
/rpt\_family="L1PA7"  
complement(16843. .17113)  
/rpt\_family="AluSp"  
complement(17203. .17655)  
/rpt\_family="MUT1C"  
20097. .20232  
/rpt\_family="(TA)n"  
complement(20777. .21168)  
/rpt\_family="LTR52"  
21905. .21935  
/rpt\_family="(TG)n"  
22022. .22154  
/rpt\_family="FLAM\_A"  
complement(23446. .23611)  
/rpt\_family="PFAM"  
23653. .23818  
/rpt\_family="(TA)n"  
23819. .23855  
/rpt\_family="(TG)n"  
23856. .23864  
/rpt\_family="(TA)n"  
complement(23865. .24165)  
/rpt\_family="AluSx"  
24375. .24584  
/rpt\_family="MIR"  
25099. .25372  
/rpt\_family="AluJo"  
25316. .25611  
/standard\_name="G08219"  
/db\_xref="dbSTS:14597"  
25373. .25407  
/rpt\_family="(TAA)n"  
25408. .25422  
repeat\_region

	repeat_region	/rpt_family="AluJo"	
	repeat_region	/complement(28866..29170)	
	repeat_region	/rpt_family="HALI"	
	repeat_region	/complement(29224..29489)	
	repeat_region	/rpt_family="HALI"	
	repeat_region	/complement(29537..29606)	
	repeat_region	/rpt_family="L1MD1"	
	repeat_region	29607..29645	
	repeat_region	/rpt_family="(TG)n"	
	repeat_region	complement(29646..29661)	
Query Match	2.7%	Score 55; DB 8;	Length 19305;
Best Local Similarity	63.0%	Pred. No. 0.11;	
Matches	85;	Conservative 0;	Mismatches 50; Indels 0; Gaps 0;
OY	4	TTTATTTGTATGAGAAAAAGAAAAGAGAAGTGAATAATTATTAACGTGCATATAGTCA	63
Db	169904	TTAATCAGGAATTTGATAATAGAAAGCATGAGTCAGATATATATGCGCTGGGCATGCAGCAG	169963
OY	64	GGACCATGATTTGCAAGTACAGAAACTCAATTCACAACCACGTAAGTCAAAAGAAAT	123
Db	169964	AGTCATATGATTTACAAAGTACAGAAAGCCAACTCAACTTGTCTTAAACAAAATGAAAT	170023
OY	124	ATATGGGCTCATGTA	138
Db	170024	TTATTTGTTTCATATA	170038
RESULT 19			
AC130404			
LOCUS	AC130404	135653 bp	DNA linear PRI 28-FEB-2003
DEFINITION	Homo sapiens 12 BAC RP11-337L12 (Rowell) Park Cancer Institute		
ACCESSION	Human BAC Library) complete sequence.		
VERSION	AC130404		
KEYWORDS	AC130404.8 GI:28436243		
SOURCE	HTG		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 135653) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbakid,J., Benton,J., Bimoge,K., Blankenburg,K., Bonini,D.P., Bouck,D., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chin,D., Chowdhury,I., Christopoulos,C., Cleaveland,C.D., Davis,C., Coyle,M.D., Dathorne,S.R., David,R., Davidson,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinib.H., Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guvera,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huliyk,S., Humel,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovac,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtearge,O., Lieu,C., Liu,J., Liu,W., Louissead,H., Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Mariondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Maxiney,E., McLeod,M.P., Meador,M., Mel.G., Mercher,S., Merker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newison,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokemwo,S., Oguh,M., Okunonu,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,		

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

1. (bases 1 to 135653)  
 Direct Submission  
 Submitted (10-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 2. (bases 1 to 135653)  
 Direct Submission  
 Submitted (07-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3. (bases 1 to 135653)  
 Direct Submission  
 Submitted (20-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 4. (bases 1 to 135653)  
 Direct Submission  
 Submitted (28-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 5. (bases 1 to 135653)  
 Direct Submission  
 Submitted (28-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Feb 20, 2003 this sequence version replaced gi:28269366.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:  
 STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.  
 Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:



http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht  
ml.

## FEATURES

source

location/Qualifiers

1..135653

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="12"

/clone="RP11-337L12"

complement(1..2004)

/note="overlaps bases 1..2004 of clone AC139767"

/function="clone overlap"

96..234

/rpt\_family="MIR"

395..419

/rpt\_family="AT-rich"

complement(721..1446)

/rpt\_family="LTR8"

1460..2018

/rpt\_family="L1P5"

2131..2164

/rpt\_family="(TA)n"

2174..2472

/rpt\_family="AluSg"

2480..2574

/rpt\_family="(TA)n"

2704..2775

/rpt\_family="L1MC5"

complement(3123..3616)

/rpt\_family="MER41B"

complement(13727..4262)

/rpt\_family="LTR49"

complement(4263..4576)

/rpt\_family="AluY"

complement(4577..4603)

/rpt\_family="LTR49"

complement(4955..5149)

/rpt\_family="MER4A2"

5150..5425

/rpt\_family="AluSc"

complement(5426..5707)

/rpt\_family="MER4A2"

complement(5731..5951)

/rpt\_family="MER58A"

complement(5952..6050)

/rpt\_family="MER113"

6051..6352

/rpt\_family="AluSg"

complement(6353..6457)

/rpt\_family="MER113"

complement(7818..7961)

/rpt\_family="L2"

complement(8067..8585)

/rpt\_family="MER1A"

9194..9224

/rpt\_family="(TG)n"

complement(9514..9676)

/rpt\_family="L2"

complement(9687..10172)

/rpt\_family="AluSx"

repeat\_region /rpt\_family="MER4D"  
complement(14783..15082)  
/rpt\_family="AluJo"

repeat\_region 18064..18242  
/rpt\_family="MIR"

repeat\_region complement(18245..18293)  
/rpt\_family="L1P4"

repeat\_region 18294..18565  
/rpt\_family="L1P4"

repeat\_region complement(19730..19832)  
/rpt\_family="L2"

Query Match 2.7%; Score 53.8; DB 8; Length 135653;  
Best Local Similarity 62.0%; Pred. No. 0.21;  
Matches 85; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 4 TTTATTTGATGAGAAAAAGAGAGTGAATAATATTACGTGATAGTTCA 63  
DB 49921 TTTCTTTGCTGATTTAAACAAAGATTGCTAATATTACATGAAATGATTGTC 49980

QY 64 GACCATGATTCAGAGAGAGAACTCAATTCAGAACTGAAAGTAAAGGAAAT 123  
DB 49981 GATATGTTATTCAGAGAGAGAACTCAATTCAGAACTGAAAGTAAAGGAAAT 50040

QY 124 ATATTGCTCATGTAAC 140  
DB 50041 TTTATGTTCTATTAAC 50057

RESULT 20  
AC023243/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 151751)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Casale,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cooke,P.,  
Dearlano,K., Dewar,K., Domino,M., Doyle,M., Feneclor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardner,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lander,T., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Margis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPherson,R., Meldrum,J., Menes,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,  
Scotjanovic,N., Sudramanian,A., Talamas,J., Teste,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A., and Zody,M.

Direct Submission  
Submitted (10-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 1, 2000 this sequence version replaced gi:6558060.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RW/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research



gap 128245..128344 /estimated\_length=100  
misc\_feature 128345..151751 /note="assembly\_fragment"

ORIGIN

Query Match 2.7%, Score 53.8; DB 14; Length 151751;  
Best Local Similarity 62.0%; Pred. No. 0.2;  
Matches 85; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 4 TTATTTGATGAGAAAAAGAGAGTGAATAATATATTAACGTGATATAGTTCA 63  
6443 TTCTTTGTCGATTAATAAAGAGATGCTATTAATTCACGAATGATATGTTGTG 6384

Qy 64 GGACCATGATGATGACATGACAACTCATTTCAACCAACGTAAGTCAAAAGAAAT 123  
6383 GATATGTTAATGACAGACAGAACCTTCAACTGGGTTAGCAAAATGGGAAAT 6324

Qy 124 ATATTGCTCATGTAAC 140  
6323 TTATTGTTCTATAAAC 6307

Db

RESULT 21  
AC041019/c 158059 bp DNA linear HTG 22-MAY-2000  
LOCUS Homo sapiens clone RP11-572N16, WORKING DRAFT SEQUENCE, 11  
DEFINITION unorderd pieces.  
AC041019  
AC041019.2 GI:8016476  
HTG: HTGS PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 158059)  
Birken, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-572N16  
Unpublished  
2 (bases 1 to 158059)  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karataa, A., Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPeckers, R., Meldrum, J., Menes, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On May 22, 2000 this sequence version replaced gi:7534221.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L9578  
Center clone name: 572\_N\_16  
----- Summary Statistics  
Sequencing vector: MJ3; M77815; 100% of reads  
Assembly: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 150828 bases at least Q40  
Consensus quality: 154584 bases at least Q30  
Consensus quality: 156053 bases at least Q20  
Insert size: 153000; agarose-fp  
Insert size: 157059; sum-of-contigs  
Quality coverage: 4.9 in Q20 bases; agarose-fp  
Quality coverage: 5.1 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
3522 3521: contig of 3521 bp in length  
\* 3522 gap of 100 bp  
\* 3622 8787: contig of 5166 bp in length  
\* 8788 8887: gap of 100 bp  
\* 8888 15129: contig of 6242 bp in length  
\* 15130 15229: gap of 100 bp  
\* 15230 23646: contig of 8417 bp in length  
\* 23647 23746: gap of 100 bp  
\* 23747 34488: contig of 10742 bp in length  
\* 34489 45971: gap of 100 bp  
\* 45972 46071: contig of 11383 bp in length  
\* 46072 61548: gap of 100 bp  
\* 61549 76890: contig of 15477 bp in length  
\* 76891 76990: gap of 100 bp  
\* 76991 93810: contig of 15242 bp in length  
\* 93811 93910: gap of 100 bp  
\* 93911 111867: contig of 16820 bp in length  
\* 111868 111957: gap of 100 bp  
\* 111958 158059: contig of 17957 bp in length  
\* 158059 46052: gap of 100 bp in length.  
Location/Qualifiers  
1. 158059  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-572N16"  
/clone\_lib="RP11-572N16 Human Male BAC"  
1. 3521  
/note="assembly\_fragment"  
3522 3621  
/note="assembly\_fragment"  
3622 8787  
/note="assembly\_fragment"  
8788 8887  
/note="assembly\_fragment"  
8888 15129  
/note="assembly\_fragment"  
15130 15229  
/note="assembly\_fragment"  
15230 23646  
/note="assembly\_fragment"  
23647 23746  
/note="assembly\_fragment"  
23747 34488  
/note="assembly\_fragment"  
34489 34588  
/note="assembly\_fragment"  
34589 45971  
/note="assembly\_fragment"



*	37075	41312	contig of 4258 bp in length
*	41313	41412	gap of 100 bp
*	41413	44561	contig of 3149 bp in length
*	44562	44661	gap of 100 bp
*	44662	49191	contig of 4530 bp in length
*	49192	49291	gap of 100 bp
*	49292	52928	contig of 3637 bp in length
*	52929	53028	gap of 100 bp
*	53029	56291	contig of 3263 bp in length
*	56292	56391	gap of 100 bp
*	56392	63001	contig of 6610 bp in length
*	63002	63101	gap of 100 bp
*	63102	66593	contig of 6492 bp in length
*	66594	66693	gap of 100 bp
*	66694	75000	contig of 5307 bp in length
*	75001	75100	gap of 100 bp
*	75101	83143	contig of 8043 bp in length
*	83144	83243	gap of 100 bp
*	83244	90745	contig of 7502 bp in length
*	90746	90845	gap of 100 bp
*	90846	98863	contig of 8018 bp in length
*	98864	98963	gap of 100 bp
*	98964	108065	contig of 9103 bp in length
*	108066	108166	gap of 100 bp
*	108167	118933	contig of 10769 bp in length
*	118936	119033	gap of 100 bp
*	119036	130633	contig of 11604 bp in length
*	130640	130739	gap of 100 bp
*	130740	144393	contig of 13653 bp in length
*	144393	144493	gap of 100 bp
*	144493	159251	contig of 14759 bp in length
*	159252	159351	gap of 100 bp
*	159352	187341	contig of 27990 bp in length
Location/Qualifiers			

```

misc_feature      1. .1003
gap               /note="assembly_fragment"
                 1004. .1103
misc_feature      /estimated_length=100
                 1104. .2892
misc_feature      /note="assembly_fragment"
gap              2893. .2992
misc_feature      /estimated_length=100
                 2993. .5079
gap              /note="assembly_fragment"
                 5080. .5179
misc_feature      /estimated_length=100
                 5180. .6911
misc_feature      /note="assembly_fragment"
gap              6912. .7011
misc_feature      /estimated_length=100
                 7012. .6719
gap              /note="assembly_fragment"
                 6720. .8819
misc_feature      /estimated_length=100
                 8820. .10691
misc_feature      /note="assembly_fragment"
gap              10692. .10791
misc_feature      /estimated_length=100
                 10792. .11292
gap              /note="assembly_fragment
                 clone_end:17
gap              vector_side:right"
                 12293. .12392
misc_feature      /estimated_length=100
                 12393. .14456
misc_feature      /note="assembly_fragment"
gap              14457. .14556

```

	misc_feature	estimated_length=100 14557. .17814 /note="assembly_fragment"
	gap	17815. .17914 /estimated_length=100 17915. .21497 /note="assembly_fragment"
	misc_feature	21498. .21597 /estimated_length=100 21598. .25074 /note="assembly_fragment"
	gap	25075. .25174 /estimated_length=100 25175. .28507 /note="assembly_fragment"
	misc_feature	28508. .28907 /estimated_length=100 28908. .32510 /note="assembly_fragment"
	gap	32511. .32710 /estimated_length=100 32711. .36954 /note="assembly_fragment"
	misc_feature	36955. .37054 /estimated_length=100 37055. .41312 /note="assembly_fragment"

Query Match	2.7%	Score 53.8;	DB 14;	Length 187341;
Best Local Similarity	62.0%	Pred. No. 0.2;		
Matches	85;	Conservative	0;	Mismatches 52;
			Indels	0;
			Gaps	0;

Oy	4	TTATTTGGTATGAGAAAAGAAAAGAGAGTGAATAATATTAACGTGCATATAGTTCA	63
Db	18999	TTCTTTTGTCTGATTTATTAACAAAGATTTGCTAAATTAATTCACGAAATGTATTAAGTTG	19058
Oy	64	GGACCAATGAGTTTCCAAATGACAAAATCTAAATTCAAACCAAGTAATGAAAAAGAAAT	123
Db	19059	GATATGTGTAATTTGCAATGACGAAACCTACTCAATACGTGGTTAGCAAAATGGGGAAAT	19118
Oy	124	ATAATTGGCTCATGTAAAC	140
Db	19119	TTATTTGGTTTCATAAAC	19135

LOCUS	AC023225	190310 bp	DNA	linear	HTG 26-MAY-2000
DEFINITION	Homo sapiens chromosome 1 clone RP11-440M16 map 1, WORKING DRAFT SEQUENCE, 18 unordered pieces.				
ACCESSION	AC023225				
VERSION	AC023225.3	GI:8076879			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 190310) Bitren,B., Linton,L., Nusbaum,C. and Lander,E.				
TITLE	Homo sapiens chromosome 1, clone RP11-440M16				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 190310) Bitren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Castle,A., Choepel,X., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeRubeis,K., Dewar,K., Domino,M., Doyle,M., Fennell,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardina,S., Grant,G., Hagos,B., Heathford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lander,E., Lehoczyk,U., Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N., McEwan,P., McGuck,A., McKernan,K.,				

TITLE  
JOURNAL  
COMMENT

McPheters, R., Meldrim, J., Neneus, L., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Srojinovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W., Z.,  
Zimmer, A. and Zody, M.

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7342073.  
All repeats were identified using RepeatMasker:  
Sult, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L6684

Center clone name: 440 M.16

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 181728 bases at least Q40

Consensus quality: 185239 bases at least Q30

Consensus quality: 166925 bases at least Q20

Insert size: 175000; agarose-fp

Insert size: 188610; sum-of-contigs

Quality coverage: 8.5 in Q20 bases; agarose-fp

Quality coverage: 7.9 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1095: contig of 1095 bp in length  
\* 1096 1195: gap of 100 bp  
\* 1196 2767: contig of 1572 bp in length  
\* 2768 2867: gap of 100 bp  
\* 2868 4870: contig of 2003 bp in length  
\* 4871 4970: gap of 100 bp  
\* 4971 6216: contig of 1246 bp in length  
\* 6217 7525: contig of 1209 bp in length  
\* 7526 7625: gap of 100 bp  
\* 7626 9452: contig of 1827 bp in length  
\* 9453 9552: gap of 100 bp  
\* 9553 11875: contig of 2323 bp in length  
\* 11876 11975: gap of 100 bp  
\* 11976 14007: contig of 2032 bp in length  
\* 14008 14107: gap of 100 bp  
\* 14108 20618: contig of 6511 bp in length  
\* 20619 20718: gap of 100 bp  
\* 20719 29063: contig of 8345 bp in length  
\* 29064 29163: gap of 100 bp  
\* 29164 36474: contig of 7311 bp in length  
\* 36475 36574: gap of 100 bp  
\* 36575 46685: contig of 10111 bp in length  
\* 46686 46785: gap of 100 bp  
\* 46786 58540: contig of 11755 bp in length  
\* 58541 58640: gap of 100 bp  
\* 58641 73486: contig of 14846 bp in length  
\* 73487 73587: gap of 100 bp  
\* 73587 89995: contig of 16409 bp in length  
\* 89996 90095: gap of 100 bp  
\* 90096 111543: contig of 21448 bp in length

# FEATURES

source 1.190310  
\* 111544 111643: gap of 100 bp  
\* 111644 143863: contig of 32220 bp in length  
\* 143864 143964 143963: gap of 100 bp  
\* 143964 190310: contig of 46347 bp in length.  
Location/Qualifiers

1.190310  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1"  
/clone="RP11-440M16"  
/clone\_lib="RPCT-11 Human Male BAC"  
1.1095  
/note="assembly\_fragment"  
misc\_feature  
1096.1195  
/estimated\_length=100  
1196.2767  
/note="assembly\_fragment"  
misc\_feature  
2768.2867  
/estimated\_length=100  
2868.4870  
/note="assembly\_fragment"  
gap  
4871.4970  
/estimated\_length=100  
4971.6216  
/note="assembly\_fragment"  
misc\_feature  
6217.6316  
/estimated\_length=100  
6317.7525  
/note="assembly\_fragment"  
gap  
7526.7625  
/estimated\_length=100  
7626.9452  
/note="assembly\_fragment"  
misc\_feature  
9453.9552  
/estimated\_length=100  
9553.11875  
/note="assembly\_fragment"  
gap  
11876.11975  
/estimated\_length=100  
11976.14007  
/note="assembly\_fragment"  
gap  
14008.14107  
/estimated\_length=100  
14108.20618  
/note="assembly\_fragment"  
gap  
20619.20718  
/estimated\_length=100  
20719.29063  
/note="assembly\_fragment"  
gap  
29064.29163  
/estimated\_length=100  
29164.36474  
/note="assembly\_fragment"  
gap  
36475.36574  
/estimated\_length=100  
36575.46685  
/note="assembly\_fragment"  
misc\_feature  
46686.46785  
/estimated\_length=100  
46786.58540  
/note="assembly\_fragment"  
gap  
58541.58640  
/estimated\_length=100  
58641.73486  
/note="assembly\_fragment"  
gap  
73487.73587  
/estimated\_length=100  
73587.89995  
/note="assembly\_fragment"  
misc\_feature  
89996.90095  
/estimated\_length=100

```
misc_feature      90096..111543
                  /note="assembly_fragment
                  clone_end:T7
                  vector_side:left"
gap              11544..111643
                  /estimated_length=100
                  11644..143863
misc_feature     143864..143963
                  /note="assembly_fragment"
                  143864..143963
                  /estimated_length=100
                  143964..150310
gap             150311..150310
misc_feature     150311..150310
                  /note="assembly_fragment
                  clone_end:SP6
                  vector_side:left"

ORIGIN
Query Match      2..7%; Score 53.8; DB 14; Length 190310;
Best Local Similarity 62.0%; Pred.No.0.2;
Matches          85; Conservative 0; Mismatches 52; Indels 0; Gaps 0;;

Oy               4 TTTATTGTGATGAGAAAAAGAAAAGAGTGCAAAATAATATTAACGTGCATATAGTTC A 63
                   |||||
Db               548 TTCTTTGTCGATATATTAACAAGAAGTGGCTAATATTAATTCACGTGAATGTATAGTGTG C 607

Oy               64 GGACCATGAGTTGCAAGTGACGAAACATTCATTCATCAACCACGTAAGTCAGCAAAAGAAAT T 123
                   |||||
Db               608 GATATGTTAATTTGCCAAGTGACGAAACCCCTACTCAAACCTGGGTGAAGCAAAATGGCGAAT G 667

Oy               124 ATATGGCTCATTGTAAAC 140
                   |||||
Db               668 TTATTGTTTCATTAAC 684

RESULT 24
AE014826        250663 bp DNA linear INV 07-OCT-2002
LOCUS           AE014826
DEFINITION      Plasmodium falciparum 3D7 chromosome 14 section 11 of 13 of the
                 complete sequence.
ACCESSION       AE014826
KEYWORDS        AE014187
SOURCE          AE014826.1 GI:23497662
ORGANISM        Plasmodium falciparum 3D7
REFERENCE       1 Plasmodium falciparum 3D7
AUTHORS         Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
                Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
                Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
                James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
                Kyes,S., Chan,M.-S., Nene,V., Shallow,S.J., Suh,B., Peterson,J.,
                Angiuoli,S., Pereira,M., Allen,J., Selengut,J., Haft,D.,
                Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
                Cunningham,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
                Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
                Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Frazer,C.M. and
                Barrell,B.
TITLE           Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL         Nature 419 (6906), 498-511 (2002)
PUBMED         12368864
REFERENCE       2 (bases 1 to 250663)
AUTHORS         Gardner,M.J.
TITLE           Direct Submission
JOURNAL         Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
FEATURES        Medical Center Dr, Rockville, MD 20850, USA
source          location/Qualifiers
                1..250663
                /organism="Plasmodium falciparum 3D7"
                /mol_type="Genomic DNA"
                /isolate="3D7"
                /db_xref="taxon:36329"
                /chromosome="14"
                complement(99..160)
                repeat_region
                /rpt_unit="(taaaaa)n"
```

repeat_region	complement (191. .258) /note="AT_rich"
repeat_region	/rpt_type=tandem complement (273. .394) /rpt_type=tandem /rpt_unit="(ta)n"
repeat_region	complement (350. .468) /rpt_type=tandem /rpt_unit="(tataa)n"
repeat_region	complement (472. .530) /note="AT_rich"
repeat_region	/rpt_type=tandem complement (576. .655) /note="AT_rich"
repeat_region	/rpt_type=tandem complement (679. .752) /rpt_type=tandem /rpt_unit="(ta)n"
repeat_region	679. .742 /rpt_type=tandem 754. .793
repeat_region	/rpt_type=tandem complement (754. .792) /rpt_type=tandem /rpt_unit="(taaaa)n"
repeat_region	complement (814. .945) /note="AT_rich"
repeat_region	/rpt_type=tandem 816. .845
repeat_region	/rpt_type=tandem 869. .94141
gene	/locus_tag="PF14_0589"
mRNA	<869. .>4141 /locus_tag="PF14_0589"
CDS	869. .4141 /locus_tag="PF14_0589"
	/codon_start=1 /product="yaline - CRNA ligase, putative"
	/protein_id="AAN37202.1"
	/db_xref="GI:23497663"
	/translation="MNIRLRKKGIGNYINKSPFYCNKUKSKYMDKNIKENITMDQAYDPKREKASISFEWENDYFKPKKELISEKNNEHKKFVITILPPVNTGTTHGHTLTVAIQDSLVRKRNKNNTLTLYTPGTDHAGIATQTVKMKLYKKENKIRQYGGEEFVKKIYEKQELHGNKNNQKRIIGASVDMGREYFATNENISNAVKEAFIFYSGSLYRNRRLVAMCPKATLSDIEVNLEBKPTKIKLPSFDHLVEGVLYKFPYQIOSEBKEIETATRIEMGLDVAVAHPKDKRYAHLIGKEIVHPEIPNKKIIIIADPFDMQYGAVKITPAHDKNDYDMKKNHNYINIPFLDGHINENGKLPQGGHREPCRKIOBELKHLNLSDKQENPMSPLCSRTNDII EYMLIPQWYNKSELAHQAIQNVKLELIIIPQOHNWYVYLLEWVRWISCSRLQMGHRLPAVKIKNDVNNTDININDNDNDNNNNNNNNLEQOHLNVVGRSYEEGLEKAKSLVGNKPEPFLIODEVLDTPFSSSLVPSSLGMNQTEDLETFEPNTIILETGODILFPWVARNVWSLHMKKLPEKTIYLAHMRIDSRGEMSSKCNVATDPLDIIDGTSINKLHKKLYEGNLPEKEIKRAIETLOKKEPKGPIECGDALAPGLLTYLKGRVNDLINRIIGYRHFCKLMAVAFPLKLTLDPNYDNRIILIDREYVEKIQWEDKMILHKLNYITNADNPESYFSSVAFASTYFWLYDLCIDYLBELIKRNLTDSDHNDVSGLEKESDMNIEKREVEQNDNNEISVNTNHEINKCDNYLDGAGNTTLTCLVDYGLRLHPISPITTEELVOKISSEPKYKNSISLAYPYINIMHNENISMKNENVIKOPRSPITTELPKTKNCPILAKHEDKFIKVPKTKLEVLAKESSSVITNVEDNSDDKLOVKKCIADITYSNOPITYVQSSSEVLYKRLSLMANKNKKIQASLIDSTLKKINDPNYEQKVPBOVRTMYSKEVELNSQILSISNIIICEINBECKNA"
repeat_region	complement (1923. .1949) /note="AT_rich"
repeat_region	/rpt_type=tandem 2320. .2410
repeat_region	/rpt_type=tandem /rpt_unit="(taa)n"
repeat_region	2367. .2403 /rpt_type=tandem complement (3950. .3970) /note="AT_rich"
repeat_region	/rpt_type=tandem complement (4187. .4256) /note="AT_rich"



```

repeat_region      /rpt_type=tandem
                    complement(4287..4349)
                    /rpt_type=tandem
                    /rpt_unit="(a)n"
repeat_region      /rpt_type=tandem
                    complement(4323..4349)
                    /rpt_type=tandem
                    /rpt_unit="(a)n"
repeat_region      /rpt_type=tandem
                    complement(4391..4413)
                    /note="AT_rich"
repeat_region      /rpt_type=tandem
                    complement(4419..4449)
                    /rpt_type=tandem
                    complement(4419..4445)
                    /rpt_type=tandem
                    /rpt_unit="(a)n"
repeat_region      /rpt_type=tandem
                    complement(4480..4562)
                    /note="AT_rich"
gene               /rpt_type=tandem
                    complement(<4588..>4953)
                    /locus_tag="PF14_0590"
mRNA               complement(<4588..>4953)
                    /locus_tag="PF14_0590"
                    complement(4588..4953)
                    /locus_tag="PF14_0590"
CDS                /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AA37203.1"
                    /db_xref="GI:23497664"
                    /translation="MKRTDRIYVDINNEEYKNI.PDDKNDLIYIIDITRMGPGICF
                    TPEMINKIYKNNLISEENYKLFSCAQYTSLSKNYDNCKPFYIIILKNEIIIOGIC
                    NLPILPSFIDEHLMKKIN"
repeat_region      complement(4781..4809)
                    /note="AT_rich"
                    /rpt_type=tandem
                    complement(4952..5000)
                    /note="AT_rich"
repeat_region      /rpt_type=tandem
                    complement(5038..5093)
                    /rpt_type=tandem
                    /rpt_unit="(a)n"
repeat_region      complement(5064..5093)
                    /rpt_type=tandem
                    complement(5094..5138)
                    /note="AT_rich"
repeat_region      /rpt_type=tandem
                    complement(5203..5243)
                    /rpt_type=tandem
                    /rpt_unit="(a)n"
repeat_region      complement(5276..5307)
                    /note="AT_rich"
repeat_region      /rpt_type=tandem
                    complement(5308..5334)
                    /rpt_type=tandem
                    /rpt_unit="(a)n"
gene               complement(<5401..>9405)
                    /locus_tag="PF14_0591"
mRNA               complement(<5401..>9405)
                    /locus_tag="PF14_0591"
                    complement(5401..9405)
                    /locus_tag="PF14_0591"
CDS                /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AA37204.1"
                    /db_xref="GI:23497665"
                    /translation="MKYIDIGLFPKSYDNHETIKYIDKLRNEIIEKENDIKEPLKESN
                    NVIDNTRIKKIYGCVEKTKKNEIDTISNKLVDNYKSTQVCNIDITYNKELNEIEK
                    DIVKNWLKKYDYELFENPPEKKNDKDEKLGNTINNIINEDSLTKESVIT
                    KYVYLINKEKKKKDVANVDCFLANNYINNIIYFDINKKLEBNYIYCIKIRYFDLT
                    SLIVLSKILNKEKIKONIYINIPEKPCNYIANNVNYKKAHPSNVIKILHYEFNLT
                    YKSCNQGTIKISIMILLIPYONKTDIINFHKKNENIKOHIDINKKONMLPHALIK
                    ARKLIHNPINKKKKCNKGNELDRSTIOKNEELIHNNFYLAENKLIHYEKKTIYVL
                    LNKTYTILLILINSCCPLLBHKKKKDKSHIYKKEICERKOPYNIDITNYKIHINK
                    ENQIDVSEYENNFLOQYKNEIFLQDIYNNKHIKNGIEKKNEFYNNINTEIDKE
                    KIINTNTVENFILLKYHYNOCIKNIYKYVLYLLANNYYTQYNIISIDPLFENYETIK

```

repeat\_region  
complement(5542..5574)  
/note="AT\_rich"

Query Match 2.6%; Score 52.8; DB 2; Length 250663;  
Best Local Similarity 52.2%; Pred. No. 0.32;  
Matches 117; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

Qy      1588 TGCTCACTATCATTAAGAAAGTCATTAGCTAAGAAACAAACTCAATCTATGTATTA 1647
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      230081 TGTTAAGTATTCATTBAAATATATATATATATATATATATATATATATAT 230140

Qy      1648 GAAGAACAAAGCTGGTTTGCCTCAATATATAAATAAAGAAACCATGTGAAGTCAA 1707
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      230141 ATGTATATAGATATATATATATATATATATATATATATATATATATATAT 230200

Qy      1708 AATATTTGTTTATCAGGTCATTGAGAAATCTATTAAAAAGTATTGGAATCTTTATGANG 1767
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      230201 TCACCGTATATATAGCTAAACATATATATGTATATATATATATATATATAT 230260

Qy      1768 AGAATATCTTGATCTCAAGTGAAGAGTGCTGAGCTTTTGTGCGCT 1811
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      230261 ATCACTATATATAGTAACACAAACAGAGAGTGACTTTGTGCACT 230304

```

RESULT 25  
AL954840  
LOCUS Zebrafish DNA sequence from clone DK91-29K19, complete sequence.  
DEFINITION AL954840  
ACCESSION AL954840  
VERSION AL954840.16 GI:37652324  
KEYWORDS HTG.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
REFERENCE Harrison,E.  
AUTHORS Submitted (11-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,  
JOURNAL Cambridgehire, CB10 1SA, UK. E-mail enquiries:  
Zfish@hdsheire.ac.uk; Clone request: clonerequest@sanger.ac.uk  
COMMENT On Oct 13, 2003 this sequence version replaced gi:37650875.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit submissions only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., paired quality >= 30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.sanger.ac.uk/projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/projects/D_rerio/fishmask.shtml) DKEY-29K19 is from a Zebrafish BAC library

## FEATURES

source

Location/Qualifiers  
1. 167004  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-29K19"  
/clone\_lib="DanioKey"

## ORIGIN

Query Match

Best Local Similarity 50.6%; Pred. No. 0.46; Length 167004;  
Matches 126; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

1547 TTGATTAGATTGGCTGAATAGACTGATCTGCCATTCTCTCATTATGATAGGA 1606  
112615 TTCTGAGACTATCGGAATACATTAAGCTTAAGGGCTAGTACCTTAACCTTAAT 112674  
1607 AGTCATTAGCTAGGAACAACCTACATCTATGTAATTGAGAACAACGCTGTTTG 1666  
112675 GGGGTTTAAAAATGTAACCTCTTTTATTTTAAAGCGAATTAACAATCGACTTTT 112734  
1667 CTCATATATAAATAAAGAAACCATGTAAGTCAAAATTTGTTTATCAGGT 1726  
112735 TCTAGATTAATAAATTTATCAACACTCTGTAATAATCCCTTGCTCTGTTAAACATTA 112794  
1727 CATTTGAAATCTTATTAAGATTTGAATCTTTATGATGAGAACTATCTGACTCAG 1786  
112795 TTTCGAAAATATTTTAAAAAGAAAAAATCTTAGGGGGCTTAATATTCGACTTCAG 112854  
1787 TGGACAGCTG 1795  
112855 CTGTGTGTG 112863

## RESULT 26

BX927412/c

188520 bp DNA linear HTG 13-MAY-2005  
Danio rerio clone CH211-143A8, WORKING DRAFT SEQUENCE, 7 unordered pieces.

## DEFINITION

BX927412

BX927412

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

McLaren, S.  
Direct Submission  
Submitted (12-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [fish-help@sanger.ac.uk](mailto:fish-help@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

## COMMENT

On May 13, 2005 this sequence version replaced gi:45822745.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [fish-help@sanger.ac.uk](mailto:fish-help@sanger.ac.uk)

----- Project Information

Center project name: ZC143A8

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 186450 bases at least Q40

Consensus quality: 186777 bases at least Q30

Consensus quality: 187184 bases at least Q20

Insert size: 187920; sum-of-contigs

Insert size: 181420; 27.5% error; agarose-fp

Quality coverage: 7.23x in Q20 bases; sum-of-contigs Quality coverage: 7.60x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 23128: contig of 23128 bp in length  
\* 23129 23228: gap of 100 bp  
\* 23229 97391: contig of 74163 bp in length  
\* 97392 97491: gap of 100 bp  
\* 97492 107937: contig of 10446 bp in length  
\* 107938 108037: gap of 100 bp  
\* 108038 167507: contig of 59470 bp in length  
\* 167508 167607: gap of 100 bp  
\* 167608 169634: contig of 2027 bp in length  
\* 169635 169734: gap of 100 bp  
\* 169735 183842: contig of 14108 bp in length  
\* 183843 183942: gap of 100 bp  
\* 183943 188520: contig of 4578 bp in length.

## FEATURES

source

1. 188520  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-143A8"  
/clone\_lib="CHORI-211"  
1. 23128  
/note="assembly fragment:00412  
fragment\_chain:1"  
23229 .97391  
/note="assembly fragment:01532  
fragment\_chain:1"  
clone\_end:SP6  
vector\_side:left"  
97492 .107937  
/note="assembly fragment:00103  
fragment\_chain:1"  
108038 .167507  
/note="assembly fragment:00696  
fragment\_chain:1"  
167608 .169634  
/note="assembly fragment:00033  
fragment\_chain:1"  
169735 .183842  
/note="assembly fragment:00226  
fragment\_chain:1"  
183943 .188520  
/note="assembly fragment:00049  
fragment\_chain:1"

## ORIGIN

Query Match

2.6%; Score 51.8; DB 14; Length 188520;

Best Local Similarity 47.2%; Pred. No. 0.55;  
Matches 158; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

OY 1434 GCTCTTAAGCGTAATGTTTATATGACTAGGCTCTGGCTTACATATATAAGGGAT 1493  
Db 108439 GGTATAGTGTATTTGTTGTTGTAAGAAATGACGTCTATTTTGGACCTATAAATAT 108380  
OY 1494 ATTGACCAATGTCATGATCAGAAAGCTTTCTCCAGGTCTCATATGTAAAGATTCATTA 1553  
Db 108379 AATGTTTATGACAGTTAATCATCTTTAGTACCAATCTGTTATCAAAAGTAAATAA 108320  
OY 1554 GATTGCTGAATAGACTGATCTGTCCTCATTTCTGCTCACTATATAGAAAGTCATT 1613  
Db 108319 AATAAAGTACTGAAACAAAAGTACTGTTAATTTTAAAGTAAAGTAAACAAAT 108260  
OY 1614 AGTAAAGCAACAAACTACATCTATGATATAGAAAGCAAGCTGGTTGCTCAATTA 1673  
Db 108259 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 108200  
OY 1674 TAAATAATAGAAAGAAACCATGTGAAAGTCAAAATATTTGTTAATCAGTCAATGAG 1733  
Db 108199 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 108140  
OY 1734 AATCTATTAATAAGTATTTGAATTTCTTTATGATGA 1768  
Db 108139 TAAATATACATTAATTAATTAATTTTGTAAATTA 108105

RESULT 27  
AC092392/c 98286 bp DNA linear PRI 15-DEC-2001  
LOCUS AC092392 Homo sapiens chromosome 1 clone RP4-544h6, complete sequence.  
DEFINITION AC092392 ALI09625  
AC092392.2 GI:17861066  
VERSION HTG.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catearrhini;  
Homidae; Homo  
REFERENCE 1 (bases 1 to 98286)  
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
Haugen, E.D.  
TITLE Direct Submission  
JOURNAL 2 (bases 1 to 98286)  
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and  
Haugen, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
REFERENCE 3 (bases 1 to 98286)  
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and  
Haugen, E.D.  
TITLE Direct Submission  
JOURNAL Submitted (15-DEC-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
COMMENT On Dec 15, 2001 this sequence version replaced gi:14589610.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: http://www.genome.washington.edu  
Contact: uwgchgs@u.washington.edu  
Drafting Center: SC  
----- Project Information  
Center Project name: Chr-1  
Center Clone name: RP4-544h6 (sc0188)  
----- Summary Statistics  
Sequencing vector: plasmid; 22% of reads  
Chemistry: Dye-terminator ETI 32% of reads  
Chemistry: Dye-terminator Big Dye; 68% of reads  
Assembly program: Phrap; version 0.990319

Consensus quality: 98273 bases at least Q40  
Consensus quality: 98278 bases at least Q30  
Consensus quality: 98286 bases at least Q20  
Insert size: 98287; sum-of-contigs  
Quality coverage: 9.1x in Q20 bases; sum-of-contigs

Overlapping Sequences:  
5': RP5-104313 (UWGC:sc0545) AC090902  
3': RP4-671122 (UWGC:sc0205) AL356298  
-----

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M3 subclone; and the assembly was confirmed by restriction digest.

#### Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BglII	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----
6023	5830	8696	8978	1849	1934	-----
-----	-----	-----	-----	-----	-----	-----
2067	2146	6101	6394	6382	6210	-----
-----	-----	-----	-----	-----	-----	-----
7084	7102	6972	6986	512	<800	-----
-----	-----	-----	-----	-----	-----	-----
2144	2146	6986	6986	449	<800	-----
-----	-----	-----	-----	-----	-----	-----
1462	1431	11036	10738	1792	1934	-----
-----	-----	-----	-----	-----	-----	-----
2877	2772	826	827	7699	7751	-----
-----	-----	-----	-----	-----	-----	-----
252	<800	1438	1380	3248	3382	-----
-----	-----	-----	-----	-----	-----	-----
547	<800	2212	2194	6225	6210	-----
-----	-----	-----	-----	-----	-----	-----
815	883	1590	1521	9	<800	-----
-----	-----	-----	-----	-----	-----	-----
308	<800	3225	3154	720	726	-----
-----	-----	-----	-----	-----	-----	-----
515	<800	3261	3312	1705	1735	-----
-----	-----	-----	-----	-----	-----	-----
1321	1320	671	<800	2231	2162	-----
-----	-----	-----	-----	-----	-----	-----
11014	10897	2171	2194	83	<800	-----
-----	-----	-----	-----	-----	-----	-----
4803	4847	3889	3911	298	<800	-----

1643	1622	2401	2382	307	<800
2140	2146	742	827	3417	3382
2925	2947	8122	8089	2700	2801
4658	4573	988	974	262	<800
3569	3631	2263	2194	1407	1374
3936	4009	825	827	5580	5488
653	<800	7300	7287	2031	2162
18	<800	1900	1898	2574	2646
1053	1111	4808	4762	2085	2162
1135	1111	2669	2653	897	907
50	<800	2712	2746	202	<800
113	<800	6452	6394	706	<800
2617	2615	2160	2194	3060	3207
689	<800	3777	3706	3662	3618
402	<800	792	<800	1702	1655
7554	7725			250	<800
592	<800			1010	1005
5487	5529			333	<800
1643	1622			233	<800
6052	6239			1223	1200
896	883			4551	4630
5326	5322			1948	1934
1036	1030			4120	3964
6626	6811			7902	8074
1378	1320			5161	5066
599	<800			543	<800
2743	2615			6163	6210

## FEATURES

```
Location/Qualifiers
1..98286
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-5446"
/clone_lib="RPCT human PAC library 4"
```

ORIGIN

	Query Match	Best Local Similarity	2.6% 73.9%	Score 51.2	DB 8	length 98286
	Matches	Conservative	0	Mismatches	23	Indels 0 Gaps 0
Oy	55	TATAGTTCAGGACATGATTCGACGACGAAACTCAATTCAACCAAGTAAGTCA				
Db	3689	TGTTTTTCAGACTCTTGACTGCTGAGTGCAGAAACTTAATTTCAACTGCATATGCAA				
Oy	115	AAGGAAATATATATGGCTCATGTAACTT				
Db	3629	AGAGAAAATGATTTGGTTCTTGTAACCTT				

RESULT	28
AC099062	
LOCUS	101840 bp DNA linear PRI 19-OCT-2002
DEFINITION	Homo sapiens chromosome 1 clone RP5-10A3L3, complete sequence.
ACCSSION	AC099062 ALI62727
VERSION	AC099062.2 GI:24137515
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 101840)  
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Seshimachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 101840)
AUTHORS	Kaul, R. K., Olson, M. V., Raymond, C. and Haugen, E. D.
TITLE	Direct Submission
JOURNAL	Submitted (08-NOV-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE	3 (bases 1 to 101840)
AUTHORS	Kaul, R. K., Olson, M. V., Zhou, Y., James, R. A., Rouse, G., Wu, Z.,

TITLE	Direct Submission
JOURNAL	Submitted (19-OCT-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT	On Oct 19, 2002 this sequence version replaced gi:16799020.

Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: <http://www.genome.washington.edu>  
Contact: uwgchgs@u.washington.edu  
Drafting Center: SC

Center clone name: RP5-1043L3 (sc0545)  
----- Summary Statistics

Sequencing vector: plasmid; 59% of reads  
Sequencing vector: plasmid; L08752; 41% of reads  
Chemistry: Dye-terminator ET; 67% of reads  
Chemistry: Dye-terminator Big Dye; 33% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 101791 bases at least Q40  
Consensus quality: 101840 bases at least Q30  
Consensus quality: 101840 bases at least Q20  
Insert size: 101840; sum-of-contigs  
Quality coverage: 8.5x in Q20 bases; sum-of-contigs

### Overlapping Sequences:

5': RP4-544H6 (UMGC:8C0188) AC0923392, 25483-bp overlap  
3': RP5-105215 AL139139, 2000-bp overlap

**Sequence Quality Assessment:**  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.



Db 21795 TGTGTTTCAGAGACTCTTGAGCTAGTGTGTCAGAAACTAATTTCAACTGCATATGCAAA 21854

QY 115 AAGGAAATATATGTCGTCAGTAACT 142

Db 21855 AGAGAAATGATTTGTTGTTCTTGACTT 21882

RESULT 29  
AE014842/c  
LOCUS

DEFINITION AE014842 253151 bp DNA linear INV 07-OCT-2002  
Plasmodium falciparum 3D7 chromosome 11 section 7 of 8 of the  
complete sequence.

ACCESSION AE014842 AE014186

VERSION AE014842.1 GI:23496321

KEYWORDS

SOURCE Plasmodium falciparum 3D7  
Plasmodium falciparum 3D7  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

ORGANISM

REFERENCE 1 (bases 1 to 253151)  
Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T.,  
James K., Bisen J.A., Rutherford K., Salzberg S.L., Craig A.,  
Kyes S., Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J.,  
Angiuoli S., Pertea M., Allen J., Selengut J., Haft D.,  
Mather M.W., Vaidya A.B., Martin D.M.A., Fairlamb A.H.,  
Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I.,  
Cummings J.M., Subramanian G.M., Mungall C., Venter J.C.,  
Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M. and  
Barrell B.

AUTHORS

TITLE Genome sequence of the human malaria parasite Plasmodium falciparum

JOURNAL Nature 419 (6906), 498-511 (2002)

PUBMED 12368864

REFERENCE 2 (bases 1 to 253151)  
Gardner M.J.  
Direct Submission  
Submitted (13-SEP-2002) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA

JOURNAL

FEATURES

source

1..253151  
/organism="Plasmodium falciparum 3D7"  
/mol\_type="genomic DNA"  
/isolate="3D7"  
/db\_xref="taxon:36329"  
/chromosome="11"  
complement(73..101)  
/note="AT rich"  
/rpt\_type=tandem  
173..279  
/rpt\_type=tandem  
/rpt\_unit="(taaaa)n"  
complement(302..328)  
/note="AT rich"  
repeat\_region  
complement(393..466)  
/rpt\_type=tandem  
/rpt\_unit="(ta)n"  
393..442  
/rpt\_type=tandem  
/rpt\_type=tandem  
440..466  
/rpt\_type=tandem  
545..578  
/rpt\_type=tandem  
complement(545..578)  
/rpt\_type=tandem  
/rpt\_unit="(ta)n"  
complement(580..673)  
/note="AT rich"  
/rpt\_type=tandem  
762..795  
/rpt\_type=tandem  
complement(762..795)  
/rpt\_type=tandem  
/rpt\_unit="(ta)n"  
repeat\_region

repeat\_region complement(869..955)  
/note="AT rich"  
/rpt\_type=tandem  
complement(956..975)  
repeat\_region  
/rpt\_type=tandem  
/rpt\_unit="(a)n"  
complement(983..1040)  
/note="AT rich"  
repeat\_region  
/rpt\_type=tandem  
1094..1121  
/rpt\_type=tandem  
/rpt\_unit="(a)n"  
complement(1149..1221)  
/note="AT rich"  
repeat\_region  
/rpt\_type=tandem  
1239..1281  
/rpt\_type=tandem  
complement(1243..1319)  
/rpt\_type=tandem  
/rpt\_unit="(ta)n"  
complement(1323..1379)  
/note="AT rich"  
repeat\_region  
complement(14621..1699)  
/note="AT rich"  
/rpt\_type=tandem  
complement(1820..1872)  
/note="AT rich"  
repeat\_region  
complement(1921..2010)  
/note="AT rich"  
/rpt\_type=tandem  
2105..2135  
/rpt\_type=tandem  
/rpt\_unit="(a)n"  
complement(2135..2176)  
/rpt\_type=tandem  
/rpt\_unit="(ta)n"  
complement(2226..2248)  
/rpt\_type=tandem  
/rpt\_unit="(taaaa)n"  
complement(2289..2362)  
/note="AT rich"  
repeat\_region  
2372..2396  
/rpt\_type=tandem  
complement(2372..2396)  
/rpt\_type=tandem  
/rpt\_unit="(a)n"  
complement(2402..2424)  
/note="AT rich"  
/rpt\_type=tandem  
complement(2438..2620)  
/note="AT rich"  
repeat\_region  
2636..2671  
/rpt\_type=tandem  
complement(2636..2671)  
/rpt\_type=tandem  
2680..2837  
/rpt\_type=tandem  
/rpt\_unit="(ataaa)n"  
complement(2763..2865)  
/rpt\_type=tandem  
/rpt\_unit="(ta)n"  
complement(2932..2954)  
/note="AT rich"  
/rpt\_type=tandem  
2970..3008  
repeat\_region

repeat__region	/rpt_type=tandem /rpt_unit="(a)n" complement (3028. .3145) /note="AT_rich"
repeat__region	/rpt_type=tandem complement (3189. .3238) /rpt_type=tandem /rpt_unit="(cataa)n" complement (3250. .3403) /rpt_type=tandem /rpt_unit="(ta)n" 3425. .3520
repeat__region	/rpt_type=tandem complement (3425. .3520) /rpt_type=tandem /rpt_unit="(a)n" 3556. .3580
repeat__region	/rpt_type=tandem /rpt_unit="(a)n" complement (3584. .3641) /note="AT_rich"
repeat__region	/rpt_type=tandem complement (3649. .3809) /rpt_type=tandem /rpt_unit="(ta)n" complement (3842. .3884) /rpt_type=tandem /rpt_unit="(a)n" 3845. .3884
repeat__region	/rpt_type=tandem 3897. .3944 /rpt_type=tandem /rpt_unit="(cataa)n" complement (3946. .3999) /note="AT_rich"
repeat__region	/rpt_type=tandem 4000. .4024 /rpt_type=tandem complement (4000. .4024) /rpt_type=tandem /rpt_unit="(a)n" complement (4026. .4104) /note="AT_rich"
repeat__region	/rpt_type=tandem 4163. .4226 /rpt_type=tandem /rpt_unit="(taaaa)n" 4232. .4267
repeat__region	/rpt_type=tandem complement (4232. .4267) /rpt_type=tandem /rpt_unit="(ta)n" complement (4339. .4359) /rpt_type=tandem /rpt_unit="(ca)n" 4380. .4407
repeat__region	/rpt_type=tandem complement (4380. .4407) /rpt_type=tandem /rpt_unit="(a)n" complement (4408. .4575) /note="AT_rich"
repeat__region	/rpt_type=tandem

	Query Match	2.6%	Score 51.2;	DB 2;	Length 253151;
	Best Local Similarity	45.78%	Pred No. 0.72;	Mismatches 253;	Indels 2; Gaps 1;
OY	Matches	215;	Conservative	0;	
Oy	1313	TTACTGCATTAATTGTTTTGAATTCAGTGTAAATTTCTACAAATTTCCCATPAAGTACTCA	1372		
Dd	1785+5	TAAAGCGATAAATAATATAAAAAATAAATAATAATATTTTTTGATACGGTTTTATATAATTTATTT	1784866		
OY	1373	CACAAGTAACCCTTCATGCCAACACTGGCTTTGGCTAAATACATCATTATATGAGAGCTG	1432		

[illegible]



----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0761E21  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP11-58N10; the clone sequenced to the right is RP11-63A11, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-761E21; actual end is at base position 77234 of RP11-63A11.

The region from 55547 to 55670 is covered only by a PCR product from clone DNA. Polymorphisms exist between AC113151 and AC11006. Data from AC113151 and AC007799 was used to finish AC11006.

#### FEATURES

##### source

```
1..123019
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="4"
  /map="4"
  /clone="RP11-761E21"
  /clone_1lb="RPCI-11"
  19..392
  /rpt_family="L2"
  393..693
  /rpt_family="Alu"
  694..923
  /rpt_family="L2"
  923..1009
  /rpt_family="L1"
  1010..1140
  /rpt_family="Alu"
  1141..1203
  /rpt_family="L1"
  1295..1600
  /rpt_family="Alu"
  1769..1896
  /rpt_family="L1"
  1923..2198
```

```
repeat_region
  /rpt_family="Alu"
  2205..2239
  /rpt_family="(GAAA)n"
repeat_region
  2650..2701
  /rpt_family="L2"
repeat_region
  3704..4984
  /rpt_family="L1"
repeat_region
  4985..5448
  /rpt_family="MER2_type"
repeat_region
  5449..5524
  /rpt_family="MER2_type"
repeat_region
  5512..5604
  /rpt_family="MER2_type"
repeat_region
  5605..6414
  /rpt_family="L1"
repeat_region
  6415..6521
  /rpt_family="MER2_type"
repeat_region
  6524..6831
  /rpt_family="Alu"
repeat_region
  6873..7093
  /rpt_family="L1"
repeat_region
  7165..7531
  /rpt_family="Mariner"
repeat_region
  7749..7861
  /rpt_family="MER1_type"
repeat_region
  8045..8197
  /rpt_family="L1"
repeat_region
  8198..8624
  /rpt_family="MALR"
repeat_region
  8625..8824
  /rpt_family="L1"
repeat_region
  9046..9327
  /rpt_family="Alu"
repeat_region
  9510..9530
  /rpt_family="AT-rich"
repeat_region
  9731..9898
  /rpt_family="MER2_type"
repeat_region
  9899..10032
  /rpt_family="MER1_type"
repeat_region
  10033..10327
  /rpt_family="Alu"
repeat_region
  10328..10408
  /rpt_family="MER1_type"
repeat_region
  10409..10582
  /rpt_family="MER2_type"
repeat_region
  10595..10908
  /rpt_family="Alu"
repeat_region
  11430..11547
  /rpt_family="Alu"
repeat_region
  11664..11753
  /rpt_family="L2"
repeat_region
  11850..13322
  /rpt_family="L1"
repeat_region
  13323..13635
  /rpt_family="Alu"
repeat_region
  13636..13666
  /rpt_family="L1"
repeat_region
  13826..14116
  /rpt_family="Alu"
repeat_region
  16039..16162
  /rpt_family="MIR"
repeat_region
  22848..22955
  /rpt_family="MER1_type"
repeat_region
  25267..25393
  /rpt_family="MIR"
repeat_region
  26158..26283
  /rpt_family="L2"
repeat_region
  26305..26353
  /rpt_family="L2"
repeat_region
  27372..27402
  /rpt_family="L2"
repeat_region
  27515..27537
  /rpt_family="(A)n"
```

	repeat_region	27897..28177	/rpt_family="Alu"
	repeat_region	29544..29564	/rpt_family="AT_rich"
	repeat_region	29566..29860	/rpt_family="Alu"
	repeat_region	30550..30673	/rpt_family="Alu"
	repeat_region	30677..30714	/rpt_family="(CAAA)n"
	repeat_region	31380..31475	
Query Match		2.5%;	Score 50.6; DB 8; Length 123019;
Best Local Similarity		65.5%;	Pred. No. 1.1;
Matches	74; Conservative	0; Mismatches	39; Indels 0; Gaps 0;
Oy	30	AGAGTGAATAATATTAACGTCATCTAGCTCAGGACCATGATTCGAAGTACAGAA	89
Db	68191	AGGAGAGATTAAAGTTTAACTTTTATTCATCTGACTCTTGTCGAAAGACAGAA	68250
Oy	90	CTCATTCGAACCAAGTAAAGTCAAAAGGAAATATTTGGCTCATGTAACTT	142
Db	68251	TCCAATCTCAACCGATTAAAGTCAAAAGGAAATTTATGGCTCATTCAACTT	68303

RESULT 31				
AC108123/c				
LOCUS	AC108123	191481 bp	DNA	linear HTG-25-JAN-2002
DEFINITION	Homo sapiens chromosome 16 clone RP11-56511, WORKING DRAFT SEQUENCE, 10 unordered pieces.			
ACCESSION	AC108123			
VERSION	AC108123..1	GI:18369969		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE AUTHORS TITLE JOURNAL RECORD AUTHORS TITLE JOURNAL COMMENT	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 191481) DOE Joint Genome Institute. Sequencing of Human Chromosome 16 unpublished 2 (bases 1 to 191481) DOE Joint Genome Institute. Direct Submission Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA -----Genome Center			

```

Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 623518
Center clone name: RPC1-11_56511
-----
Summary Statistics
Consensus quality: 182477 bases at least Q40
Consensus quality: 187876 bases at least Q30
Consensus quality: 188750 bases at least Q20
Estimated insert size: 192000; agarose-tp estimation
Estimated insert size: 190581; sum-of-contigs estimation
Quality coverage: 6.71 in Q20 bases; agarose-tp estimation
Quality coverage: 6.76 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1230: contig of 1230 bp in length

```

FEATURES	source
*	1231 1330: gap of unknown length
*	1331 3606: contig of 2276 bp in length
*	3607 3706: gap of unknown length
*	3707 6459: contig of 2753 bp in length
*	6460 6559: gap of unknown length
*	6560 10289: contig of 3730 bp in length
*	10290 10389: gap of unknown length
*	10390 16864: contig of 6475 bp in length
*	16865 16964: gap of unknown length
*	16965 27807: contig of 10843 bp in length
*	27808 27907: gap of unknown length
*	27908 39098: contig of 11191 bp in length
*	39099 39198: gap of unknown length
*	39199 76534: contig of 37336 bp in length
*	76535 121347: gap of unknown length
*	121348 121447: contig of 44713 bp in length
*	121448 191481: gap of unknown length
	191481 191481: contig of 70034 bp in length
	Location/Qualifiers
	1..191481

gap	1231	1330	/clone_1lib-RPCI human BAC library 11
gap	3607	3706	/estimated_length=unknown
gap	6460	6559	/estimated_length=unknown
gap	10290	10389	/estimated_length=unknown
gap	16865	16864	/estimated_length=unknown
gap	27808	27907	/estimated_length=unknown
gap	39099	39198	/estimated_length=unknown
gap	76535	76634	/estimated_length=unknown
gap	121448	121447	/estimated_length=unknown

[illegible]

KEYWORDS	JP 2001269182-A/15200.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 431)
TITLE	Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.
JOURNAL	Sequence tag and encoded human protein
COMMENT	Patent: JP 2001269182-A 15200 02-OCT-2001; GENSET
OS	Homo sapiens (human)
PN	JP 2001269182-A/15200
PD	02-OCT-2001
PF	24-FEB-2000 JP 2000118773
PR	26-FEB-1999 US 60/122487
PI	JEAN BAPTISTE DUMAS MILNE EDWARDS, FIMERIC DUCLAIR, JEAN YVES JORDAN
PC	C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC C12N5/10,
CC	PC C12P21/02, C12P21/08, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, G06F15/40
FEATURES	Location/Qualifiers.
SOURCE	1..431
ORIGIN	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
Query Match	2.5%; Score 49.4; DB 6; Length 431;
Best Local Similarity	71.4%; Pred. No. 3.9;
Matches	65; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Oy	70 TGGATTGCAAGTGACAGAACTCAATTCAAACCAAGTAAGTCAAAAGAAATATATTG 129
Db	231 TTGGTTGTAAGTGAACAGAAACCCCACTCAAAATTAGCTTAGGCAAAAAGAGAAATATATGG 290
Oy	130 GCTCATGTAACCTTCTCAAGAGAGGGCAGG 160
Db	291 GTTCATGAAACCCCAACTTAAGAAAGGACG 321
RESULT 33	
AX903421	431 bp DNA linear PAT 18-DEC-2003
LOCUS	AX903421
DEFINITION	Sequence 19284 from Patent EPI033401.
ACCESSION	AX903421
VERSION	AX903421.1 GI:40058378
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
REFERENCE	1
AUTHORS	Dumas Milne Edwards, J.B., Duclair, A. and Giordano, J.Y.
TITLE	Expressed sequence tags and encoded human proteins
JOURNAL	Genet. EP 1033401-A 19284 06-SEP-2000;
FEATURES	Location/Qualifiers
SOURCE	1..431
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
ORIGIN	
Query Match	2.5%; Score 49.4; DB 6; Length 431;
Best Local Similarity	71.4%; Pred. No. 3.9;
Matches	65; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Oy	70 TGGATTGCAAGTGACAGAACTCAATTCAAACCAAGTAAGTCAAAAGAAATATATTG 129

[illegible]



REFERENCE 1 (bases 1 to 157671)  
AUTHORS Babbage, A.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk  
On Mar 5, 2001 this sequence version was replaced gi:11318483.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
Eml; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP13-225021 is from the library RPc1-13.1 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
Location/Qualifiers  
1. 157671  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP13-225021"  
/clone\_1fb="RPC1-13.1"  
1  
/note="Clone\_left\_end: RP13-225021"  
1749..1858  
/note="Sequence from overlapping clones BA342E24 (AL357623) and BA131B16 (AL354705). Assembly confirmed by restriction digest."  
join(AL160169.12:31066..31267,AL160169.12:78517..78638,AL160169.12:82136..82243,AL160169.12:93987..94037,AL356862.10:28698..28781,AL356862.10:44310..44399,AL356862.10:47114..47206,AL356862.10:63497..63623,AL357623.14:16855..16992,AL357623.14:19465..19558,AL357623.14:25501..25568,8456..8582,9839..9996,12380..12431,25496..25601,25992..26084,26309..26414,28116..28207,32105..36529)  
/gene="RP13-225021.1"  
/locus\_tag="RP13-225021.1-001"  
join(AL160169.12:31066..31267,AL160169.12:78517..78638,AL160169.12:82136..82243,AL160169.12:93987..94037,AL356862.10:28698..28781,AL356862.10:44310..44399,AL356862.10:47114..47206,AL356862.10:63497..63623,AL357623.14:16855..16992,AL357623.14:19465..19558,AL357623.14:25501..25568,8456..8582,9839..9996,12380..12431,25496..25601,25992..26084,26309..26414,28116..28207,32105..36529)  
/gene="RP13-225021.1"  
/locus\_tag="RP13-225021.1-001"  
/note="match: ESTs: AA121710.1 AA126485.1 AA126611.1 AA284968.1 AA293632.1 AA354061.1 AA446069.1 AA455215.1 AA478123.1 AA478986.1 AA815347.1 A1183582.1 A1582306.1 A1613219.1 A1656043.1 A1697716.1 A1909150.1 A1909967.1 AL600005.1 AU119994.1 AU134174.1 AV753402.1 AW118605.1 AW137800.1 AW138212.1 AW750651.1 AW842484.1 AW959164.1

AW968411.1 AW976742.1 BP196253.1 BP951871.1 BG433408.1  
BG984586.1 BI050462.1 B1598980.1 B1916008.1 BM505956.1  
BM681237.1 BM685093.1 BM726588.1 BM741025.1 BM745776.1  
BM746243.1 BM753571.1 BM764434.1 BM765703.1 BM766713.1  
BM766542.1 BM772138.1 BM772477.1 BM932077.1 BM983757.1  
BM023410.1 BQ165492.1 BQ188845.1 BQ218851.1 BQ228362.1  
BQ324015.1 CA392005.1 CA412508.1 F07861.1 H12126.1  
H18534.1 H19388.1 H55229.1 W28684.1 W28685.1 W28701.1  
match: cDNAs: AB002349.1"  
join(AL160169.12:78582..78638,AL160169.12:82136..82243,AL160169.12:93987..94037,AL356862.10:28698..28781,AL356862.10:44310..44399,AL356862.10:47114..47206,AL356862.10:63497..63623,AL357623.14:16855..16992,AL357623.14:19465..19558,AL357623.14:25501..25568,8456..8582,9839..9996,12380..12431,25496..25601,25992..26084,26309..26414,28116..28207,32105..32134)  
/gene="RP13-225021.1"  
/locus\_tag="RP13-225021.1-001"  
/standard\_name="ORTHUMP00000022161"  
/note="match: proteins: O15059"  
/codon\_start=1  
/protein\_id="CA141461.1"  
/db\_xref="GI:57209377"  
/db\_xref="GOA:O5VXM4"  
/db\_xref="InterPro:IPR001849"  
/db\_xref="UniProt/TREMBL:O5VXM4"  
/translation="MTKNGLMASVIVTSATQGGSSDLSLGGSCDVAASKYDAVF  
DVLKTPLEFAQILMDIPVKAIQPEBLACGMSKSKSLAPVAVAFTRFQVS  
FVWVIEILAQILKIRAEILSHVIAIKLNLMSVLSVLSAQSAPIRLTFTW  
ALNKRDKTFEKLPIYLMKEDNYKREYISILKWPISPIGLIIDLITDYSAP  
AGSISIMENORNONNMLIRIADIQVSCSRLITLPIVOKVILSVRTIELOKVE  
DNYKLSLRIFEGSSSPRVSKEDLAGSASGARRRRTCPDPTSVAGSLPTTPV  
PBRKSHSLGNMWCQLSVESKSAIFPEGRKHLDDVLSRSPRLALITSSAV  
TNGLSIGSESESEFSEBMSGLSPGPCICLGSNAVPTVEGRLRRRTLLKEGKP  
ALSSWTIRVWVILSGSTLYYGAKSLRGTDRIKYEKTPGKYSIVGMVLPDPPEHPD  
FOLNPNPDGVNYKFOGTSRFPALHMKHLDDPACSNRQOVANLMSFE"  
join(AL160169.12:78517..78638,AL160169.12:82136..82243,AL160169.12:93987..94037,AL356862.10:28698..28781,AL356862.10:44310..44399,AL356862.10:47114..47206,AL356862.10:63497..63623,AL357623.14:16855..16992,AL357623.14:19465..19558,AL357623.14:25501..25568,8456..8582,24060..24143,25496..25601,25992..26084,26333..26414,28116..28207,30271..31177)  
/gene="RP13-225021.1"  
/locus\_tag="RP13-225021.1-002"  
join(AL160169.12:78517..78638,AL160169.12:82136..82243,AL160169.12:93987..94037,AL356862.10:28698..28781,AL356862.10:44310..44399,AL356862.10:47114..47206,AL356862.10:63497..63623,AL357623.14:16855..16992,AL357623.14:19465..19558,AL357623.14:25501..25568,8456..8582,24060..24143,25496..25601,25992..26084,26333..26414,28116..28207,30271..31177)  
/gene="RP13-225021.1"  
/locus\_tag="RP13-225021.1-002"  
/note="match: ESTs: AA121487.1 AA478789.1 A1627675.1 AW245646.1 AW250005.1 BE263432.1 BP510295.1 CA392004.1  
match: cDNAs: AF221098.1"  
join(AL160169.12:78582..78638,AL160169.12:82136..82243,AL160169.12:93987..94037,AL356862.10:28698..28781,AL356862.10:44310..44399,AL356862.10:47114..47206,AL356862.10:63497..63623,AL357623.14:16855..16992,AL357623.14:19465..19558,AL357623.14:25501..25568,8456..8582,24060..24143,25496..25601,25992..26084,26333..26414,28116..28207,30271..30366)  
/gene="RP13-225021.1"  
/locus\_tag="RP13-225021.1-002"  
/standard\_name="ORTHUMP00000022162"  
/note="match: proteins: O9NZ16"  
/codon\_start=1  
/protein\_id="CA141460.1"  
/db\_xref="GI:57209376"  
/db\_xref="GOA:O9NZ16"

```
/db_xref="InterPro:IPR001849"
/db_xref="InterPro:IPR001895"
/db_xref="UniProt:TrEMBL:O9N216"
/translation="MYKRNLMASVLTATPGSSSSDLSLGGSCDVASKSYDAVVF
DUKATPEEPASQITTLMDIPVFAKIOPEELASGMSKKEKSHIAPVPAFTRRNQVS
FWVRILTAQTLKIAELISHFVKIAKLELHSLMSVSAQSAPIFRLITVM
ALNRDQKTFEKLIDYLMSEKEDYKRTREYRSLKNVPSIPYGLYLDLILYDSXP
AGSGIMENQORSQNMNILLIADLVSCYDHLTLPYQKYLKSVRYIELOKFE
DNYKSLRIEPPGSSSPRLVSKEEDLAGSGASRFSRPFPCPTSVAGSLPFPV
PBRKSHSLGNNGRLYATLGPWRYVPVNSPPTBSCVSPGPCIISGNSAAPTVM
EGPLRKTLKEGRKPAKLSWTRWYILSGSTLLYYGASLRKTDKSHSYIYWMYOL
PDDPEHPDIFQLNPPKGNVYKFTGSRFHAILMHRKLDLACKSNRPOBGAAPGPTG
TDSHEVDHLEGAGKAGPCA"
misc_feature
17109..17357
/gene="RP13-225021.1"
/locus_tag="RP13-225021.1-002"
/note="Sequence from overlapping clones ba342824
(AL357623) and ba131B16 (AL354705). Assembly confirmed by
restriction digest."
join(23868..23890,24060..24143,25496..25601,25992..26084,
26309..26332,26385..26397)
/gene="RP13-225021.1"
/locus_tag="RP13-225021.1-003"
join(23868..23890,24060..24143,25496..25601,25992..26084,
26309..26332,26385..26397)
/gene="RP13-225021.1"
/locus_tag="RP13-225021.1-003"
join(24110..24143,25496..25601,25992..26084,26309..26414,
28116..28207,32105..36531)
/gene="RP13-225021.1"
/locus_tag="RP13-225021.1-005"
join(24110..24143,25496..25601,25992..26084,26309..26414,
28116..28207,32105..36531)
/gene="RP13-225021.1-005"
/locus_tag="RP13-225021.1-005"
/note="match: cDNAs: BC037725.1"
CDS
join(<24110..24143,25496..25601,25992..26084,26309..26414,
28116..28207,32105..36531)
/gene="RP13-225021.1-005"
/locus_tag="RP13-225021.1-005"
/note="match: cDNAs: BC037725.1"
Query Match 2.5%; Score 49.4; DB 8; Length 157671;
Best Local Similarity 71.4%; Pred. No. 1.9;
Matches 65; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 70 TGGATTGCAAGTACGAGAACTCAATCAACCACTAAGTCAAAAGAAATATATG 129
60652 TTGGTTGTAGTACGAGAAACCAACTCAATTAAGTACGCAAAAAGAAATATATG 60711
Db 130 GCTCATGTAACTTCTTCACAGAGAGGCGAG 160
60712 GTTCATGGAACCAACTAAGAAAGGAGCG 60742
RESULT 37
AL354705/c 161027 bp DNA linear HTG 16-OCT-2000
LOCUS Homo sapiens chromosome 9 clone RP11-131B16, WORKING DRAFT
DEFINITION SQUONCE, 3 unordered pieces.
ACCESSION AL354705
VERSION AL354705.10 GI:10862755
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1
REFERENCE
1 Plumb,B.
AUTHORS Direct Submission
TITLE Submitted (15-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT requests: clonequest@sanger.ac.uk
On Oct 17, 2000 this sequence version replaced gi:10834465.
----- Genome Center
```

```
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba131B16
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 160247 bases at least Q40
Consensus quality: 160580 bases at least Q30
Consensus quality: 160708 bases at least Q20
Insert size: 160827; sum-of-coverage
Insert size: 16550; agarose-gel
Quality coverage: 6.63x in Q20 bases; sum-of-coverage
Quality coverage: 6.63x in Q20 bases; agarose-gel
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 67522: contig of 67522 bp in length
* * 67523 67622: gap of 100 bp
* * 67623 104112: contig of 36490 bp in length
* * 104113 104212: gap of 100 bp
* * 104213 161027: contig of 56815 bp in length.
FEATURES
source
1..161027
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-131B16"
/clone_1fb="RPC1-11.1"
1..67522
/note="assembly fragment:00751
fragment chain:1
clone end:17
vector side:left"
67623..104112
/note="assembly fragment:01022
fragment chain:1"
104213..161027
/note="assembly fragment:02056
fragment chain:1
clone end:SP6
vector_side:right"
ORIGIN
Query Match 2.5%; Score 49.4; DB 14; Length 161027;
Best Local Similarity 71.4%; Pred. No. 1.9;
Matches 65; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 70 TGGATTGCAAGTACGAGAACTCAATCAACCACTAAGTCAAAAGAAATATATG 129
81740 TTGGTTGTAGTACGAGAAACCAACTCAATTAAGTACGCAAAAAGAAATATATG 81681
Db 130 GCTCATGTAACTTCTTCACAGAGAGGCGAG 160
81680 GTTCATGGAACCAACTAAGAAAGGAGCG 81650
RESULT 38
PFMAL4P3/c 349751 bp DNA linear INV 16-APR-2005
LOCUS Plasmodium falciparum MAL4P3.
DEFINITION
ACCESSION AF035476
VERSION AL035476.7 GI:23498195
```

KEYWORDS Plasmodium falciparum 3D7  
SOURCE Plasmodium falciparum 3D7  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1  
AUTHORS Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D., Mungall, K., Bowman, S., Ackin, R., Baker, S., Barron, A., Brooks, K., Buckee, C.O., Burrows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Cotton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J., Felwell, T., Godle, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsbey, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K.D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P., Moulé, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M.A., Rabinowletsch, E., Rajandream, M.A., Rutter, S., Rutherford, K.M., Sanders, C., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J.E., Craig, A., Newbold, C. and Barrell, B.G.  
TITLE Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13  
JOURNAL Nature 419 (6906), 527-531 (2002)  
PUBMED 12368867  
REFERENCE 2 (bases 1 to 349751)  
AUTHORS Hamlin, N., Pain, A., Berriman, B., Hall, N., Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M. and Barrell, B.  
TITLE Direct Submission  
JOURNAL Submitted (24-FEB-1999) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
COMMENT On Oct 3, 2002 this sequence version replaced gi:5763808. For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).  
FEATURES  
source  
1. 349751  
/organism="Plasmodium falciparum 3D7"  
/mol\_type="genomic DNA"  
/isolate="3D7"  
/db\_xref="taxon:36329"  
/chromosome="4"  
/clone="MAL4P3"  
/gene="PF0630C"  
/complement(join(2130..3551,4472..9988))  
/note="synonym: VAR"  
/complement(join(2130..3551,4472..9988))  
/gene="PF0630C"  
/note="Similar to Plasmodium falciparum var, mal4p2.58 mal4p2.58, var SMALL:Q9UG05 (EMBL:AL035475) (2277 aa) and to Plasmodium falciparum variant-specific surface protein var-7 SMALL:Q26034 (EMBL:U42636) (2182 aa) faasta scores: E(): 1.9e-126, 46.72% id in 2412 aa"  
/codon\_start=1  
/product="erythrocyte membrane protein 1 (pFEMP1)"  
/protein\_id="CAD49167.1"  
/db\_xref="GI:23498197"  
/db\_xref="GOA:Q8I1V7"  
/db\_xref="InterPro:IPR004258"  
/translation="MGPPSTAPDPYSAKHALDSIGKRVHAOVONKAKORSNDLKGILTSATLSGSGEAFTEPNCFLIKKRGDLSDSGRHPCGNSASERKFSVSGDECDKKTIRGNRNGGACAPYRRLPLCNKNEKKEKRTSTGHLLDLCANAYEASLIRYHDTQDATYPSGDSFMCWTMLARSPADIDGIRGDLVGRKKKKYTEROQSLSKIKTIGDIYELNENGRGVNDHYODDNGVYFOLREDMVANAATWKAITCAADTGNAYKPTCSNQGQSAHHCRCNGDKDDPKDPTDPTDYVVOYLRFWEBAEAFGRCKKIIYGVITKICREKYSNGPRYCSRNGYDCTKTKRAIGKRYMNGQICISLYACNPYVDMINOKEQFDKOKKYYTDEINERSSRSRQKRGARSTGSSSNVYDYESPYNLIKDYGYVDDFLKLINKEKSCQAVKNDGCTINTEKDDKONNNKDKGTYSREYCYQVDCGVATNSGWEKKKNDQCNIKLYRKKAAPHTIKLKGEGKEETLEKLEAEACDKDGGNSDLSLVDPMQCYQFDLEKDEKEGVDRANDVDTGGGLCILQKLEBENKCKOKTYNDPFNVAHMLKDSIHWKKLEKCIQNGTKTRCNNEKCNKCECFGRWAKKIKEMWKIKDHFPTQNGIDETNCDPIVTLRGVLKGLPLNEDSTQDKNSLDSSELKILK

misc\_feature  
13697..13898  
/note="GC-rich region"  
gene  
complement(join(17269..18690,19611..25169))  
/gene="PF0635C"  
/note="synonym: VAR"  
/complement(join(17269..18690,19611..25169))  
/note="PF0635C"  
/note="Similar to Plasmodium falciparum var, mal4p2.58 mal4p2.58, var SMALL:Q9UG05 (EMBL:AL035475) (2277 aa) and to Plasmodium falciparum variant-specific surface protein var-7 SMALL:Q26034 (EMBL:U42636) (2182 aa) faasta scores: E(): 7.9e-126, 46.25% id in 2441 aa"  
/codon\_start=1  
/product="erythrocyte membrane protein 1 (pFEMP1)"  
/protein\_id="CAD49167.1"  
/db\_xref="GI:23498197"  
/db\_xref="GOA:Q8I1V6"  
/db\_xref="InterPro:IPR004258"  
/translation="NAPTSGGGTDESKHALDSIGKRVHAOVONKAKORSNDLKGILTSATLSGSGEAFTEPNCFLIKKRGDLSDSGRHPCGNSASERKFSVSGDECDKKTIRGNRNGGACAPYRRLPLCNKNEKKEKRTSTGHLLDLCANAYEASLIRYHDTQDATYPSGDSFMCWTMLARSPADIDGIRGDLVGRKKKKYTEROQSLSKIKTIGDIYELNENGRGVNDHYODDNGVYFOLREDMVANAATWKAITCAADTGNAYKPTCSNQGQSAHHCRCNGDKDDPKDPTDPTDYVVOYLRFWEBAEAFGRCKKIIYGVITKICREKYSNGPRYCSRNGYDCTKTKRAIGKRYMNGQICISLYACNPYVDMINOKEQFDKOKKYYTDEINERSSRSRQKRGARSTGSSSNVYDYESPYNLIKDYGYVDDFLKLINKEKSCQAVKNDGCTINTEKDDKONNNKDKGTYSREYCYQVDCGVATNSGWEKKKNDQCNIKLYRKKAAPHTIKLKGEGKEETLEKLEAEACDKDGGNSDLSLVDPMQCYQFDLEKDEKEGVDRANDVDTGGGLCILQKLEBENKCKOKTYNDPFNVAHMLKDSIHWKKLEKCIQNGTKTRCNNEKCNKCECFGRWAKKIKEMWKIKDHFPTQNGIDETNCDPIVTLRGVLKGLPLNEDSTQDKNSLDSSELKILK



ANDDNKINFPKREKTFNPKCKAKCPVYGVKTYGGNYIPNEBKDYKSKKGVKEND  
TIPNIEVLVIGRKEGKEDKXHLHDACKNTGLFEVARYEONMCKKGIIDOCKITKE  
ANDIDBDKDIIVNEBFQRMRLRFVODYNNLKKKINCTKKEKEKEKXKCOGND  
KCECYEMKSKKQEWTOIKTYKXYSKI SDEIARVYVSVDOLFPNDYKKAQEV  
VERPCDKEKXKCTGDNLKEGEDPGKCHMGDFITNLISLQKIDCKNKAONSVET  
QSDENATACQCDTHPDDEEDLLEENENVOAOPNCPNVEDKKEEVEKETAQYT  
AESTATYAGGERQTPAPAPAPAPPPRPPKPPPPDPLPALNMLSTIMWSGI  
GFAAFYFPLKKKTAKVGNLFQILOIPKGDYDIPLKSSPKRYTIPASDRYKTYIY  
MGSDSGDEKVAFMSDPTDVTSESEYFELDVNDIYVPSPKRYKTLIEVLEPQNT  
PDSIDSDNTPTTPQPTTDDENMOLKEDFISNMLOQNTOPNPIHNNVNNHPTMR  
FNNMQKPTFMSIHDRNLFSGEETNDMFNSGNPIHSDTNSMDSLTNSHSPINDK  
NDLYSGIDILINALSGNMIKIDEMLRKRENELGQHPKNTSNRVVOTQSSDPI  
HNOLNLFHTMLDRHDMCEKIKNDNERLAKLEEMENHSGINSIGNHVLTDI  
VSIQIHDNDPKPINEFTMDTSPKSTMTDITIDLEKXNEPYDIEYEDIDYDVND  
DKTSMNNNNLVKNNKPNVDSNSTVHNHNPADINKTFVDIINNHNQPIEKPKIQIEM  
NSNRREVDEQVPIADICVNTI"  
/complement (join(28486, .29547, 29708, .29776))  
/gene="PPD0640c"  
/note="synonym: RIF"  
/complement (join(28486, .29547, 29708, .29776))  
/gene="PPD0640c"  
/note="similar to: Plasmodium falciparum rifin pfcl115w",  
mal3p7.54 SMALL:O97311 (EMBL:AL034559) (371 aa) fasta  
scores: E(): 4.8e-58, 49.73% id in 382 aa, and to  
Plasmodium falciparum rifin pf0035C SMALL:O96113  
(EMBL:AE001367) (375 aa) fasta scores: E(): 1.9e-46,  
43.52% id in 386 aa"  
/codon\_start=1  
/product="RIFIN"  
/protein\_id="CAD49168.1"  
/db\_xref="GI:23498198"  
/db\_xref="InterPro:IPR002858"  
/db\_xref="InterPro:IPR006373"  
/db\_xref="UniProt/TrEMBL:O611V5"  
/translation="MKVHTISLIFSHPNILITSSQVYNQKRPYTPHKMKPKSVKT  
CRSLCEELYAPSNYNDPEMKVKQKQFDRISQSHSEHKRIQENRQCKQCEQVDA  
OKIILDKIEKELTEKLGALQTEIRSDAIPTCQSVGKQKCTLCKGGLIGESVA  
PSLGLIGELGALVINNMKATPFEYAFVAFQKEGIAAGKIASPAADIVISGIIINF  
EVHTINGSTLANITLLETLKDDITLTKALHFEGSGMVCNTPDDKICIVGMRAGVIO  
GKASPEAVIRRSVKTLKKNADVAQAQAAQTANERTSGMIKELSKIASAGANTYSA  
ITYSTAILVIVLVMVYIYILIRRRKKKKKLOQYIKLKE"  
/join(34475, 34549, 34788, .35753)  
/gene="synonym: RIF"  
/note="synonym: RIF"  
/join(34475, 34549, 34788, .35753)  
/gene="PPD0645w"  
/note="Signal peptide predicted for mal1p4.07 by SignalP  
2.0 HMM (Signal peptide probability 0.895, signal anchor  
probability 0.008) with cleavage site probability 0.389  
between residues 24 and 25"

Query Match 2.5%; Score 49.4; DB 2; Length 349751;  
Best Local Similarity 45.2%; Pred. No. 1.7; Indels 0; Gaps 0;  
Matches 182; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

DB 1309 AATTTTACGCAATATATTTGAAATCAGCATTAATTTTCAATATTTCCCAATGCA 1368  
302480 AAAAAAT 302421

DB 1369 TCTACACACAATACCTCTCATGCAACACTGGCTTGTCAATATCATATCTATATGAGA 1428  
302420 TACAATGACA 302361

DB 1429 GCTGTGCTTCTTAAGGTAATGTTTATATGACACTAAGGCTGTGGCTTACATATATATA 1488  
302360 TTTTGTATCT 302301

DB 1489 GGGGATATGACATATGATATACAGAGTCTTTCTCACAGGCTCATATGTAAGAAAT 1548  
302300 ATGATTTTCAAAATTTTAAAGAAATACATATTTTACAAAAAGAACAAATTTGTTATCATG 302241

DB 1549 CATTAGATTGGCGAAATAGACTGATCTGTCCATTTCTGTCTCATCTTATCATAGAGAG 1608  
|||||

DB 302240 GATTGTATATATATATGAGAAATATATATTTAACAAGTTCACAAAAAATATATATAT 302181

QY 1609 TCATTAGCTTAAGGACAAAAACTATCATCTATGTAATATAGAAACAAGCTGTTTGGCT 1668  
|||||

DB 302180 ATAAAAAT 302121  
|||||

QY 1669 CATTAT 1711  
|||||

DB 302120 CACAAAAAATATGACAT 302078  
|||||

RESULT 39  
CR405684/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

CR405684  
155357 bp DNA linear HTG 21-MAR-2005  
Dario relio clone DKEX-186117, WORKING DRAFT SEQUENCE, 4 unordered  
pieces  
CR405684  
GI:61673441  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Dario relio (zebrafish)  
Dario relio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 155357)  
Submitted (18-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Mar 21, 2005 this sequence version replaced gi:47058533.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zki186117  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 154173 bases at least Q40  
Consensus quality: 154392 bases at least Q30  
Consensus quality: 154589 bases at least Q20  
Insert size: 155057; sum-of-contrigs  
Insert size: 160173; 1.4% error; agarose-fp  
Quality coverage: 8.48x in Q20 bases; sum-of-contrigs Quality  
coverage: 8.21x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contrigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contrigs are represented as  
\* runs of 'N', but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1 39024: contrig of 39024 bp in length  
\* 39025 39124: gap of 100 bp  
\* 39125 96064: contrig of 56940 bp in length  
\* 96065 96164: gap of 100 bp  
\* 96165 109642: contrig of 13478 bp in length  
\* 109643 109742: gap of 100 bp  
\* 109743 155357: contrig of 45615 bp in length.  
Location/Qualifiers  
1.155357  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone\_1ib="DanioKey"  
1.39024  
/note="assembly\_fragment:00157"

```

fragment Chain:1"
misc_feature      39125..96064
                  /note="assembly_fragment:01230
                  fragment_chain:1"
misc_feature      96165..109642
                  /note="assembly_fragment:00004"
misc_feature      109743..155357
                  /note="assembly_fragment:00650.0"
ORIGIN

```

```

Query Match      2.5%; Score 49.2; DB 14; Length 155357;
Best Local Similarity 51.4%; Pred. No. 2.1;
Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

```

```

QY      1535 ATATGTAAAGAAATTCATTAGATGGCTGAATAGACATGCTGCTCTGCTGCTAC 1594
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      110081 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 110022

QY      1595 TTATCATAGAGAGTCAATAGTCAAGACAAACATCAATCATATGTAATTAAGAGAC 1654
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      110021 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 109962

QY      1655 AACCTGCTTTGCTCATATATATATATATATATATATATATATATATATATATAT 1714
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      109961 AAAAAATATTTATATATATATATATATATATATATATATATATATATATATAT 109902

QY      1715 GTTTAATCAGCTCATTGAGAAATCTATTTAAAGATTGTAAT 1756
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      109901 AATTTATTTTATATATATATATATATATATATATATATATATATATATATAT 109860

```

```

RESULT 40
CR354428/c      214395 bp      DNA      linear      VRT 27-JUN-2005
LOCUS           Zebrafish DNA sequence from clone DKEY-20D16 in linkage group 6,
DEFINITION      complete sequence.
ACCESSION      CR354428
VERSION        CR354428.16 GI:68262651
KEYWORDS       HTG.
SOURCE         Danio rerio (zebrafish)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                Cypriniformes; Cyprinidae; Danio.
                1 (bases 1 to 214395)
                Auger,K.
REFERENCE      Direct Submission
AUTHORS       Submitted (25-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
TITLE         Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
JOURNAL        zfish-help@sanger.ac.uk Clone requests:
                http://www.sanger.ac.uk/Projects/D_rerio/fags.shtml#dataeight
                On Jun 27, 2005 this sequence version replaced GI:66710852.
                ----- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: zfish-help@sanger.ac.uk

```

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

DKEY-20D16 is from a Zebrafish BAC library

VECTOR: pindigobac-5

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. Any regions longer than 1kb tagged as misc-feature 'unsure' are part of a tandem repeat of more than 10kb in length where it has not been possible to anchor the base differences between repeat copies. The region has been built up based on the repeat element to match the total size of repeat indicated by restriction digest, but repeat copies may not be in the correct order and the usual finishing criteria may not apply.

```

FEATURES
Source
Location/Qualifiers
1..214395
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="6"
/clone="DKEY-20D16"
/clone_lib="DanioKey"

```

## ORIGIN

```

Query Match      2.4%; Score 49; DB 5; Length 214395;
Best Local Similarity 47.3%; Pred. No. 2.2;
Matches 148; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

```

```

QY      1537 ATGTAAAGAAATTCATTAGATGGCTGAATAGACATGCTGCTCTGCTGCTACT 1596
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      5245 ATGGAAGAACTCATTTATTTGGCAAGGTGATGCTTAATTTAGAGTTCAGCAAAAA 5186

QY      1597 ATCATAGAGAGTCAATAGCTAAGACAAACATCAATCTATGTAATTAAGACAA 1656
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      5185 ATAAAAATATATATATATATATATATATATATATATATATATATATATATATAT 5126

QY      1657 GCTGCTTTGCTCATATATATATATATATATATATATATATATATATATATAT 1716
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      5125 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 5066

QY      1717 TTATCATAGCTCATTGAGAAATCTATTTAAAGATTGTAATTTGATGAGAACTATC 1776
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      5065 TTTTCTTTTAAAGAGAGAAAAAAGTAACGTAGATTAATTTATATATATATCTT 5006

QY      1777 TTGACTCAAGTGACAGCTGCTGCTTTTGGCTGCTGCTCCTACGTAGAAAGAGCT 1836
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      5005 TTATTAAGACAACTTAATGATGTTTCTGAACTGGAAGTCAAAAAATGAATTAACCT 4946

QY      1837 TTGTCATTAAGTC 1849
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      4945 TTGAGCAACAGTC 4933

```

```

RESULT 41
AC087447/c      113695 bp      DNA      linear      PRI 17-JUL-2002
LOCUS           Homo sapiens chromosome 11, clone RP11-655D7, complete sequence.
DEFINITION      AC087447
ACCESSION      AC087447
VERSION        AC087447.5 GI:21886886
KEYWORDS       HTG.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Homiidae; Homo.
                1 (bases 1 to 113695)
REFERENCE      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
AUTHORS       Barra,N., Bastien,V., Boguslavskiy,L., Boukhaltier,B., Brown,A.,
TITLE         Camarata,U., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
JOURNAL        Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
                Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
                Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
                Hages,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
                REFERENCES 1 to 113695)

```

Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Martinis, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

3 (bases 1 to 113695)  
Birtten, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Submitted (07-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

4 (bases 1 to 113695)  
Birtten, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Submitted (17-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 17, 2002 this sequence version replaced g1:21702864.

## COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

## Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WtBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11851

Center clone name: 655\_D\_7

## FEATURES

source

-----  
Only the last 113.7 kb of this clone are being submitted.  
The remainder overlaps accession number AC010768 [WICGR project L2144].

Location/Qualifiers

1. .113695

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone="RP11-655D7"

/clone\_11b="RP11 Human Male BAC"

30. .329

/rpt\_family="AluSg"

complement(432. .558)

/rpt\_family="MIR"

complement(1389. .1496)

/rpt\_family="MIR"

complement(1761. .1936)

/rpt\_family="L1ME3A"

1945. .1972

/rpt\_family="AT-rich"

complement(3265. .3466)

/rpt\_family="MIR"

4091. .4207

/rpt\_family="CT-rich"

complement(4216. .4525)

/rpt\_family="Aluub"

4645. .4752

/rpt\_family="MIR"

4754. .4861

/rpt\_family="CATATAA"

complement(4979. .5082)

/rpt\_family="MIR"

5779. .5845

/rpt\_family="MER113"

6272. .6302

/rpt\_family="TTTA"

6303. .6326

/rpt\_family="AT-rich"

complement(6327. .11887)

/rpt\_family="L1PA7"

complement(11889. .12180)

/rpt\_family="AluSx"

complement(12994. .13169)

/rpt\_family="MIR"

complement(13271. .13368)

/rpt\_family="L2"

13389. .13786

/rpt\_family="MLT1U"

13867. .14161

/rpt\_family="AluSx"

14265. .14569

/rpt\_family="AluSc"

complement(14598. .14709)

/rpt\_family="MIR"

complement(14918. .15517)

/rpt\_family="L2"

complement(15678. .15736)

/rpt\_family="MIR3"

16898. .17273

/rpt\_family="MLT1B"

17551. .18001

/rpt\_family="L1MC4a"

18011. .18187

/rpt\_family="L1MC4"

18382. .18509

/rpt\_family="MIR"

complement(18517. .18824)

/rpt\_family="Aluub"

19081. .19157

repeat\_region

repeat\_region

repeat\_region

```
/rpt_family="MIR"
complement(19165..19335)
/rpt_family="MLT11"
complement(19362..19484)
/rpt_family="MLT11"
19489..19645
/rpt_family="MIR"
19850..19883
/rpt_family=" (TTTTTC)n"
complement(19962..20006)
/rpt_family="L2"
20913..21209
/rpt_family="AluSp"
21213..21318
/rpt_family="MIR"
complement(21562..21731)
/rpt_family="L2"
21987..22134
/rpt_family="L3"
23078..23145
/rpt_family="A-rich"
complement(23579..23766)
/rpt_family="MIR"
complement(23824..23974)
/rpt_family="FRAM"
complement(23994..24149)
/rpt_family="MIR"
complement(24468..24524)
/rpt_family="L1MB4"
complement(24686..24925)
/rpt_family="MER20"
24948..25387
repeat_region
```

Query Match 2.4%; Score 48.4; DB 8; Length 113695;  
Best Local Similarity 71.1%; Pred. No. 3.3;  
Matches 64; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
QY 50 GTCATATATCTCAGCAGCATGATTCAGTGCAGCAAACTCAATCAACCACTGTA 109
Db 79588 GACACGAGAGGATTCCTCTTGGTTCAGTGCAGCAAACTCAATCAACCACTGTA 79529
QY 110 GTCAAAAGAAATATATGCTCATGTA 139
Db 79528 GCAAAAAGAAATATATGCTCTCTGTA 79499
```

RESULT 42  
AC021170/c 176697 bp DNA linear HTG 07-JUL-2000  
LOCUS Homo sapiens chromosome 11 clone RP11-655D7, WORKING DRAFT

AC021170  
SEQUENCE: 24 unordered pieces.  
AC021170.4 GI:7235310  
HTG: HTGS PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens

REFERENCE  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 176697)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT  
On Mar 13, 2000 this sequence version replaced gi:7024164.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center

Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0655D07  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: plasmid; 0%  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 164861 bases at least Q40  
Consensus quality: 168822 bases at least Q30  
Consensus quality: 170901 bases at least Q20  
Insert size: 194000; agarose-fp  
Insert size: 174397; sum-of-coverage  
Quality coverage: 3.57 in Q20 bases; agarose-fp  
Quality coverage: 4.02 in Q20 bases; sum-of-coverage  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1 1086: contig of 1086 bp in length  
\* 1087 1186: gap of unknown length  
\* 1187 4671: contig of 3485 bp in length  
\* 4672 4771: gap of unknown length  
\* 4772 7902: contig of 3131 bp in length  
\* 7903 8002: gap of unknown length  
\* 8003 11020: contig of 3018 bp in length  
\* 11021 11120: gap of unknown length  
\* 11121 14825: contig of 3705 bp in length  
\* 14826 14925: gap of unknown length  
\* 14926 17464: contig of 2539 bp in length  
\* 17465 17564: gap of unknown length  
\* 17565 21883: contig of 4319 bp in length  
\* 21884 21983: gap of unknown length  
\* 21984 25335: contig of 3352 bp in length  
\* 25336 25435: gap of unknown length  
\* 25436 28938: contig of 3503 bp in length  
\* 28939 29038: gap of unknown length  
\* 29039 32405: contig of 3367 bp in length  
\* 32406 32505: gap of unknown length  
\* 32506 36284: contig of 3779 bp in length  
\* 36285 36384: gap of unknown length  
\* 36385 39520: contig of 3136 bp in length  
\* 39521 39620: gap of unknown length  
\* 39621 43560: contig of 3940 bp in length  
\* 43561 43660: gap of unknown length  
\* 43661 48581: contig of 4921 bp in length  
\* 48582 48681: gap of unknown length  
\* 48682 52884: contig of 4203 bp in length  
\* 52885 52984: gap of unknown length  
\* 52985 57434: contig of 4450 bp in length  
\* 57435 57534: gap of unknown length  
\* 57535 62596: contig of 5062 bp in length  
\* 62597 62696: gap of unknown length  
\* 62697 69408: contig of 6712 bp in length  
\* 69409 69508: gap of unknown length  
\* 69509 76036: contig of 6528 bp in length  
\* 76037 76136: gap of unknown length  
\* 76137 83650: contig of 7514 bp in length  
\* 83651 83750: gap of unknown length  
\* 83751 93283: contig of 9533 bp in length  
\* 93284 93383: gap of unknown length  
\* 93384 102351: contig of 8968 bp in length  
\* 102352 102451: gap of unknown length  
\* 102452 129254: contig of 26803 bp in length  
\* 129255 129354: gap of unknown length  
\* 129355 176697: contig of 47343 bp in length.

FEATURES  
source  
Location/Qualifiers  
1. .176697  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="RP11-655D7"  
1087. .1186  
/estimated\_length=unknown  
4672. .4771  
/estimated\_length=unknown  
7903. .8002  
/estimated\_length=unknown  
11021. .11120  
/estimated\_length=unknown  
14826. .14925  
/estimated\_length=unknown  
17465. .17564  
/estimated\_length=unknown  
21884. .21983  
/estimated\_length=unknown  
25336. .25435  
/estimated\_length=unknown  
28939. .29038  
/estimated\_length=unknown  
32406. .32505  
/estimated\_length=unknown  
36285. .36384  
/estimated\_length=unknown  
39521. .39620  
/estimated\_length=unknown  
43561. .43660  
/estimated\_length=unknown  
48582. .48681  
/estimated\_length=unknown  
52885. .52984  
/estimated\_length=unknown  
57435. .57534  
/estimated\_length=unknown  
62597. .62696  
/estimated\_length=unknown  
69409. .69508  
/estimated\_length=unknown  
76037. .76136  
/estimated\_length=unknown  
83651. .83750  
/estimated\_length=unknown  
93284. .93383  
/estimated\_length=unknown  
102352. .102451  
/estimated\_length=unknown  
129255. .129354  
/estimated\_length=unknown  
ORIGIN  
Query Match 2.4%; Score 48.4; DB 14; Length 176697;  
Best Local Similarity 71.1%; Pred. No. 3.1;  
Matches 64; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 50 GTGCATTAAGTTCAGACATGATTCAGAGTGAAGAACTCAATTCAAACCAACGTAA 109  
DB 154947 GAGCAGGAGGTATTCCTCTTGTTGTCAGTGCAGAGAACTTAATGAAGAACTGCGTTAA 154888  
QY 110 GTCAAAAGAAATATATTGGCTCANGTAA 139  
DB 154887 GCNAAAAGAAATATATTGGCTCTGTAA 154888  
RESULT 43  
AC019143/C AC019143 190737 bp DNA linear HTG 07-JUL-2000  
LOCUS Homo sapiens chromosome 11 clone RP11-706A13, WORKING DRAFT  
DEFINITION SEQUENCE, 28 unordered pieces.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
AC019143  
AC019143.3 GI:7715665  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
Waterson, R.H.  
1 (bases 1 to 190737)  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 190737)  
Waterson, R.H.  
Direct Submission  
Submitted (30-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On May 7, 2000 this sequence version replaced gi:7024180.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H.NH0706A13  
----- Summary Statistics -----  
Sequencing vector: M13; 78k  
Chemistry: Dye-primer ET; 78% of reads  
Chemistry: Dye-terminator Big Dye; 22% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 167325 bases at least Q40  
Consensus quality: 175317 bases at least Q30  
Consensus quality: 180039 bases at least Q20  
Insert size: 193000; agarose-fp  
Insert size: 188037; sum-of-contigs  
Quality coverage: 3.04 in Q20 bases; sum-of-contigs  
Quality coverage: 3.18 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1325: contig of 1325 bp in length  
\* 1326 1425: gap of unknown length  
\* 1426 2586: contig of 1161 bp in length  
\* 2587 2686: gap of unknown length  
\* 2687 4338: contig of 1652 bp in length  
\* 4339 4438: gap of unknown length  
\* 4439 5989: contig of 1551 bp in length  
\* 5990 6089: gap of unknown length  
\* 6090 8269: contig of 2180 bp in length  
\* 8270 8369: gap of unknown length  
\* 8370 10239: contig of 1870 bp in length  
\* 10240 10339: gap of unknown length  
\* 10340 13484: contig of 3145 bp in length  
\* 13485 13585: gap of unknown length  
\* 13585 18186: contig of 4602 bp in length  
\* 18187 18286: gap of unknown length  
\* 18287 23764: contig of 5478 bp in length  
\* 23765 23864: gap of unknown length  
\* 23865 27992: contig of 4128 bp in length  
\* 27993 31755: contig of 3663 bp in length  
\* 31756 31855: gap of unknown length  
\* 31856 36229: contig of 4374 bp in length  
\* 36230 36329: gap of unknown length  
\* 36330 40157: contig of 3828 bp in length

*	40158	40257:	gap of unknown length
*	40258	45494:	contig of 5237 bp in length
*	45495	45594:	gap of unknown length
*	45595	51920:	contig of 6326 bp in length
*	51921	52020:	gap of unknown length
*	52021	58428:	contig of 6406 bp in length
*	58427	58526:	gap of unknown length
*	58527	64892:	contig of 6366 bp in length
*	64893	64992:	gap of unknown length
*	64993	72930:	contig of 7938 bp in length
*	72931	73030:	gap of unknown length
*	73031	81807:	contig of 8777 bp in length
*	81808	81907:	gap of unknown length
*	81908	89406:	contig of 7499 bp in length
*	89407	89506:	gap of unknown length
*	89507	99182:	contig of 9676 bp in length
*	99183	99283:	gap of unknown length
*	99283	111934:	contig of 12552 bp in length
*	111835	111934:	gap of unknown length
*	111935	126258:	contig of 14324 bp in length
*	126259	126358:	gap of unknown length
*	126359	134920:	contig of 8562 bp in length
*	134921	135020:	gap of unknown length
*	135021	145335:	contig of 10315 bp in length
*	145336	145435:	gap of unknown length
*	145436	162083:	contig of 16648 bp in length
*	162084	162183:	gap of unknown length
*	162184	175617:	contig of 13434 bp in length
*	175618	175717:	gap of unknown length
*	175718	190731:	contig of 15020 bp in length

FEATURES	SOURCE
----------	--------

misc_feature	1. .1335
gap	/note="assembly_name:Contig4"
	1326. .1425
	/estimated_length=unknown
misc_feature	1426. .2586
gap	/note="assembly_name:Contig5"
	2587. .2686
	/estimated_length=unknown
misc_feature	2687. .4338
gap	/note="assembly_name:Contig6"
	4339. .4438
	/estimated_length=unknown
misc_feature	4439. .5985
gap	/note="assembly_name:Contig7"
	5990. .6089
	/estimated_length=unknown
misc_feature	6090. .8269
gap	/note="assembly_name:Contig8"
	8270. .8369
	/estimated_length=unknown
misc_feature	8370. .10239
gap	/note="assembly_name:Contig9"
	10240. .10339
	/estimated_length=unknown
misc_feature	10340. .13484
gap	/note="assembly_name:Contig10"
	13485. .13584
	/estimated_length=unknown
misc_feature	13585. .18186
gap	/note="assembly_name:Contig11"
	18187. .18286
	/estimated_length=unknown
misc_feature	18287. .23764
gap	/note="assembly_name:Contig12"
	23765. .23864
	/estimated_length=unknown

	misc_feature	2385.	.27992	/note="assembly_name:Contig13"	
	gap	2793.	.28092	/estimated_length=unknown	
	misc_feature	28093.	.31755	/note="assembly_name:Contig14"	
	gap	31756.	.31855	/estimated_length=unknown	
	misc_feature	31856.	.36229	/note="assembly_name:Contig15"	
	gap	36230.	.36329	/estimated_length=unknown	
	misc_feature	36330.	.40157	/note="assembly_name:Contig16"	
	gap	40158.	.40257	/estimated_length=unknown	
	misc_feature	40258.	.45194	/note="assembly_name:Contig17"	
	gap	45195.	.45594	/estimated_length=unknown	
	misc_feature	45595.	.51920	/note="assembly_name:Contig18"	
	gap	51921.	.52020	/estimated_length=unknown	
	misc_feature	52021.	.58426	/note="assembly_name:Contig19"	
	gap	58427.	.58526	/estimated_length=unknown	
	misc_feature	58527.	.64892	/note="assembly_name:Contig21"	
	gap	64893.	.64992	/estimated_length=unknown	
	misc_feature	64993.	.72930	/note="assembly_name:Contig22"	
	gap	72931.	.73030	/estimated_length=unknown	
	misc_feature	73031.	.81807	/note="assembly_name:Contig23"	
	gap	81808.	.81907	/estimated_length=unknown	
	misc_feature	81908.	.89106	/note="assembly_name:Contig24	
	gap	89407.	.89506	/estimated_length=unknown	
	misc_feature	89507.	.99182	/note="assembly_name:Contig25"	
	gap	99183.	.99282	/estimated_length=unknown	
	misc_feature	99283.	.111834		
	Query Match	2.4%	Score 48.4;	DB 14;	Length 190737;
	Best Local Similarity	71.1%;	Pred. No. 3.1;	Matches 64;	Conservative 0; Mismatches 26; Indels 0; Gaps 0;
OY	50 GTGCATATAGTTCACGACCAGATGTGAAGGACAGAACTCAATTCAAACCAACGTAA	109			
Db	128302 GAGCAGGGAGGATTAATCCCTCTTGTTGCAAGGCGAAGAACTAATGAAAACCTGGCTTAA	128243			
OY	110 GTCAAAAGAAAATATATTGGCTCATGTAA	139			
Db	128242 GCATAAAGAAAATATATTGGCTCTGTAA	128213			
RESULT 44					
LOCUS	AC090163	194642 bp	DNA	linear	PRI 30-AUG-2002
DEFINITION	Homo sapiens chromosome 11, clone RP11-706A13, complete sequence.				
ACCESSION	AC090163				
VERSION	AC090163.10	GI:22549876			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				





Query Match	2.4%	Score 48.4	DB 8	Length 196642
Best Local Similarity	71.1%	Pred. No. 3.1		
Matches 64	Conservative 0	Mismatches 26	Indels 0	Gaps 0
QY	50	GTGATATATAGTTCAGGACCATGATGTTCCAAAGTGACGAAACTAATTCACCAAGCTAA	109	
Db	44544	GAGAGGAGGATATCTCCTCTTGTTGTCACAAAGTGACGAAACTTAATGAAAACTGGCTTAA	44485	
QY	110	GTCAAAAGGAAATATATATTTGGCTCATGTAA	139	
Db	44484	GCAAAAGGAAATATATATTTGGCTCTCTTAA	44455	

	ORGANISM	Candidatus Carsonella ruddii; Bacteria; Proteobacteria; Gammaproteobacteria; Candidatus Carsonella.
REFERENCE	1 (bases 1 to 1818)	
AUTHORS	Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.	
TITLE	Phylogenetic analysis of vertically transmitted psyllid endosymbionts (Candidatus Carsonella ruddii) based on atpAcd and rpoC: comparisons with 16S-23S rDNA-derived phylogeny	
JOURNAL	Curr. Microbiol. 42 (6), 419-421 (2001)	
PUBMED	11381334	
AUTHORS	2 (bases 1 to 1818)	
TITLE	Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.	
JOURNAL	Direct Submission Submitted (12-May-2000) Section of Microbiology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA	
FEATURES	Location/Qualifiers	
source	1..1818	
	/organism="Candidatus Carsonella ruddii"	
	/mol_type="genomic DNA"	
	/specific_host="Trioxa urticae"	
	/db_xref="taxon:114186"	
gene	<1..656	
	/gene="atpA"	
	<1..656	
	/gene="atpG"	
	/codon_start=3	
	/transl_table=11	
	/product="ATP synthase alpha subunit"	
	/protein_id="AAK55931.1"	
	/db_xref="GI:14193361"	
	/translation="VFYLHRLRLRSGSKVNFLEKKNI FNKNGTLTAFLIETLEG DVSIFPTNVAISITDQILPLDLM,FNSGRIPA INVGISVRVGSAGQYIKVKLSGD I RIMLAQYRELSESSSDLDIETKQLVNGEKITLTMCKRPKNYSIELVLILLII KNYFPFKIPVEKEIEFEFENINIVIRMKKEIKIKIDININLEYLIEINFPIGSL"	
gene	553..1405	
	/gene="atpG"	
	653..1405	
	/gene="atpG"	
	/codon_start=1	
	/transl_table=11	
	/product="ATP synthase gamma subunit"	
	/protein_id="AAK55932.1"	
	/db_xref="GI:14193362"	
	/translation="MNIRDIKCKINILFNINKLTNTMSGISPSKMKLIKOCPLNLGL YLRHKLIPELVNPKKNNFCCLILFTNGSGNSISEIKYCIKFLKTNKNDLLII GKRAUWEPNKNNILITKKILFDREKOIPENSELKITLNVENYVPLSSIKPNNNIKII KNLVEYKIKGNPFYEIKRNINYIKLINNYINFTLKLYLBENTFSBLKLMTTMSATDSS KLIIKNNNTIKNKIRPFKVTOEMLEIINSINL"	
gene	1402..>1818	
	/gene="atpD"	
	1402..>1818	
	/gene="atpD"	
	/codon_start=1	
	/transl_table=11	
	/product="ATP synthase beta subunit"	
	/protein_id="AAK55933.1"	
	/db_xref="GI:14193363"	
	/translation="MTRGRIIQIGLSILDVEFFKKNP KTYNALPIEYVNVYLEVOOQI GKNI VRTAIGANTGMKRINSVIDTGEPMTFPGDDCTLGILLIVLDEPIDNKGNLNFI KKPRIHLSPPKSFSEOIFNNNILLETGIKIIDLCPLPK"	
ORIGIN		
Query Match	2.4%:	Score 48.2; DB 1; Length 1818;
Best Local Similarity	45.9%:	Pred. No. 6;
Matches 164; Conservative	0;	Mismatches 193; Indels 0; Gaps 0;
OY	1415	TATCTATTAVGAGCGTGTCTTTAAAGCGTAAGTGGTTTATATGCACTAAGCTCTTG 1474
Dd	472	TATTAAATTAAAAAAGTTATTATTTAAAAAAAATACACAGTTAAAGAAATTGAATTTTTC 531
OY	1475	GCTTACATATAAAGGGGATTTGACGCAATGTATACAGAAGCTTTTTCACAGAGCTTC 1534
Dd	532	AATTATATATAATTATATATATATAGAAATVTGAAATTTGAAAAAATATATATAATTTGATA 591





C	93	40.6	2.0	6794	6	ABK31265	Abk31265 signal tr
C	94	40.6	2.0	6794	6	ABL70220	Ab170220 Chemical
C	95	40.6	2.0	6794	6	AAS61175	Aas61175 Human gen
C	96	40.6	2.0	24939	6	ABL70569	Ab170569 Chemical
C	97	40.6	2.0	110000	6	ABA92787_0	Ab92787 Buchnera
C	98	40.4	2.0	15373	6	ABL32467	Ab132467 Human imm
C	99	40.4	2.0	57130	8	ABQ77243	Abq77243 Human MAR
C	100	40.2	2.0	433	14	ADZ70956	Adz70956 Human chr

ALIGNMENTS

RESULT 1

AAV83939 standard; DNA; 80595 BP.

AAV83939,

03-MAR-1999 (first entry)

HC-contig derived from normal human chromosome 10q25.2 region.

Yeast artificial chromosome; YAC; probe; eukaryotic chromosome; neocentromere; replication; extra-chromosomal element; segregation; cell division; artificial chromosome; gene therapy; mardel(10); human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.

Homo sapiens.

W09851790-A1.

19-NOV-1998.

13-MAY-1998; 98MO-AU000352.

13-MAY-1997; 97AU-00006784.

26-AUG-1997; 97AU-00008791.

(AMRA-) AMRAD OPERATIONS PTY LTD.

Choo K, Du Sart D, Cancellia MR;

WPI; 1999-009773/01.

New isolated nucleic acid comprising neocentromere sequences from eukaryotic chromosome - used to produce replicable, segregating artificial chromosomes that can carry large amounts of DNA for gene therapy.

Claim 8; Fig 6; 540pp; English.

The present sequence represents the HC-contig derived from normal human chromosome 10, 10q25.2 region. This region can be naturally mutated to produce an unusual chromosomal marker referred to as mardel(10). The mardel(10) marker is mitotically stable and contains a functional neocentromere at a location regarded as non-centromeric. This neocentromere maps to q25.2 on chromosome 10. The specification describes nucleic acid sequences derived from a eukaryotic chromosome, including a neocentromere or its functional derivative or hybrid, that are able, in a compatible cell, of replicating, acting as extra-chromosomal element and segregating during cell division. The sequences can be used to construct artificial chromosomes for use in gene therapy comprising a replicable, segregating nucleic acid that confers a specific phenotype on cells. Human artificial chromosomes can propagate in human cells and carry large amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal, they are not mutagenic. The artificial chromosomes are also useful for generation of transgenic plants and animals, in production of proteins and to make diagnostic reagents, e.g. for expression of cytokines, receptors and growth factors, or to increase the copy number of a gene in a cell. The constructs may also be used for functional and structural analysis of chromosomes

SQL	Sequence	80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;
Query Match	100.0%;	Score 2001; DB 2; Length 80595;
Best Local Similarity	100.0%;	Pred. No. 0;
Matches 2001; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	ATCTTATTTTGTATGAGAAAAAGAGAGTGAATAATATTAACGTGCATATAGT 60
DB	78595	ATCTTATTTTGTATGAGAAAAAGAGAGTGAATAATATTAACGTGCATATAGT 78654
QY	61	TCAGAGCCATGATTTGCAGATGACAGAACTCAATTCACCAAGCTAATAAGGAA 120
DB	78655	TCAGAGCCATGATTTGCAGATGACAGAACTCAATTCACCAAGCTAATAAGGAA 78714
QY	121	AATATATTGGCTCATATGTAACCTTCTCAGAGAGGGCAGATGAGAGGGCTTTGGSAAC 180
DB	78715	AATATATTGGCTCATATGTAACCTTCTCAGAGAGGGCAGATGAGAGGGCTTTGGSAAC 78774
QY	181	AAGAGAAATTTGTTCTCAAAATTTCTAGGAATATCTAGATTAAGTCCAGATGGGTCACTTCTCT 240
DB	78775	AAGAGAAATTTGTTCTCAAAATTTCTAGGAATATCTAGATTAAGTCCAGATGGGTCACTTCTCT 78834
QY	241	GTCCCTGAGAGTGTGTGATGCCGATGGTATAGCTTTATGAGAGGAAGAGTGCATGTTAGGA 300
DB	78835	GTCCCTGAGAGTGTGTGATGCCGATGGTATAGCTTTATGAGAGGAAGAGTGCATGTTAGGA 78894
QY	301	TGAAGGTAGGGCTTAAGCAAAACAGGGCAAGGGCCACTATATCATCTTAATAATGGTTTTT 360
DB	78895	TGAAGGTAGGGCTTAAGCAAAACAGGGCAAGGGCCACTATATCATCTTAATAATGGTTTTT 78954
QY	361	TTTGATGTCTTCTCTTAATTTTCAAAATGCTTCCAAACAAATGACACACAGAAAAAGAAC 420
DB	78955	TTTGATGTCTTCTCTTAATTTTCAAAATGCTTCCAAACAAATGACACACAGAAAAAGAAC 79014
QY	421	ATAGGACCTCTACTGCTGGTGTCTTTTATCTTAACCTTTGATCTTCTTTTACAGCTTA 480
DB	79015	ATAGGACCTCTACTGCTGGTGTCTTTTATCTTAACCTTTGATCTTCTTTTACAGCTTA 79074
QY	481	CTCACTGCTTTGACCTGAGAGCCATATGCTTAAAGGCTTCTGACAGGGTTCTTCTAAG 540
DB	79075	CTCACTGCTTTGACCTGAGAGCCATATGCTTAAAGGCTTCTGACAGGGTTCTTCTAAG 79134
QY	541	CTGGGTTCTTATATATGCTCTCTCCATTTCTGTTGCTCACTCTAGATGATCTTCTCTT 600
DB	79135	CTGGGTTCTTATATATGCTCTCTCCATTTCTGTTGCTCACTCTAGATGATCTTCTCTT 79194
QY	601	TTTCTCACTCTGAGGACATGATGCTGTTTATATGACATGCTTACTTTGCTTTGGGTTT 660
DB	79195	TTTCTCACTCTGAGGACATGATGCTGTTTATATGACATGCTTACTTTGCTTTGGGTTT 79254
QY	661	TTTCTGAGGACATATGCTTCTAGATTTATCTAGACCAATAATACTACAGCCACTGGGCCA 720
DB	79255	TTTCTGAGGACATATGCTTCTAGATTTATCTAGACCAATAATACTAGCCACTGGGCCA 79314
QY	721	GGCTTCTCTCTCAATGACATGATGTTCCAGGGCTTTCACCTTATGTTAGTCAAG 780
DB	79315	GGCTTCTCTCTCAATGACATGATGTTCCAGGGCTTTCACCTTATGTTAGTCAAG 79374
QY	781	CATTCTTGCAAAAGAAAGGCTTACTTAACATATGACATTTCTAGCATTTGATTTCTTTTG 840
DB	79375	CATTCTTGCAAAAGAAAGGCTTACTTAACATATGACATTTCTAGCATTTGATTTCTTTTG 79434
QY	841	ACATGTTGTAAGATCTAATTCACATTTTGTATTAAGAATTCCTCATGGAACCAACAC 900
DB	79435	ACATGTTGTAAGATCTAATTCACATTTTGTATTAAGAATTCCTCATGGAACCAACAC 79494
QY	901	GAACTAAGCTGTCTCTGGAATGCAAGGTGGCTCTCTCAATACAGATGTTCTAGAGAGCT 960
DB	79495	GAACTAAGCTGTCTCTGGAATGCAAGGTGGCTCTCTCAATACAGATGTTCTAGAGAGCT 79554
QY	961	GTAATTTTGGGCACTTAATTTCTCACTTAAGGGGACAGACACTGAATAATTAACACAC 1020
DB	79555	GTAATTTTGGGCACTTAATTTCTCACTTAAGGGGACAGACACTGAATAATTAACACAC 79614

Oy	1021	TAACTTGTGATGTCATCGATAGTAATGCTGAGGACGACGACCTCAGAGGTGGAACCTGAC	1080
Db	79615	TAACTTGTGATGTCATCGATAGTAATGCTGAGGACGACGACCTCAGAGGTGGAACCTGAC	79674
Oy	1081	CTCTTATGTGTGTCACAGCCTTTCTTCCTTCAGAAAGTCAGCTGTGTTTTCTGCTGACTCTC	1140
Db	79675	CTCTTATGTGTGTCACAGCCTTTCTTCCTTCAGAAAGTCAGCTGTGTTTTCTGCTGACTCTC	79734
Oy	1141	CATGGAACATCAGTCTTGAAATCCTCAGACCAACATCTGGAGTATGTAATGCTCCTGACA	1200
Db	79735	CATGGAACATCAGTCTTGAAATCCTCAGACCAACATCTGGAGTATGTAATGCTCCTGACA	79794
Oy	1201	GTCCTAAGATGTCCTACCGCTGGATCTCAAAAGCGTGTGACACACGGGAGAGAGAAT	1260
Db	79795	GTCCTAAGATGTCCTACCGCTGGATCTCAAAAGCGTGTGACACACGGGAGAGAGAAT	79854
Oy	1261	GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAAGCCCTTAATTTTACTGCA	1320
Db	79855	GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTTCACAAAGCCCTTAATTTTACTGCA	79914
Oy	1321	TAAATATTTTGAATTCAGTATATTTCTACAAATTTTCCATATGATCTACACAAAT	1380
Db	79915	TAAATATTTTGAATTCAGTATATTTCTACAAATTTTCCATATGATCTACACAAAT	79974
Oy	1381	ACCCTCTCAGCAACATTTGGCTTGTATATACATATCTATATGAGCTGTGCTCTT	1440
Db	79975	ACCCTCTCAGCAACATTTGGCTTGTATATACATATCTATATGAGCTGTGCTCTT	80034
Oy	1441	AAGGTAAATGTTTATATATGCACTAAGGCTCTTGCTTACATATAAAGGGGTATTGAC	1500
Db	80035	AAGGTAAATGTTTATATATGCACTAAGGCTCTTGCTTACATATAAAGGGGTATTGAC	80094
Oy	1501	AATGTGATACAGAAAGTCTTTTCTCACAAGCTCTCATATGTAAGAAATTCATTATGATGGC	1560
Db	80095	AATGTGATACAGAAAGTCTTTTCTCACAAGCTCTCATATGTAAGAAATTCATTATGATGGC	80154
Oy	1561	TGAAATGAGCTGATCTGTCAATTTCTGTGCTCACTTATCATAGAAAGTCATTAGCTAG	1620
Db	80155	TGAAATGAGCTGATCTGTCAATTTCTGTGCTCACTTATCATAGAAAGTCATTAGCTAG	80214
Oy	1621	GAAACAATACTACAACTATATATTAATTAAGAAACAAGCTGTTTTGCTCAATATAAAT	1680
Db	80215	GAAACAATACTACAACTATATATTAATTAAGAAACAAGCTGTTTTGCTCAATATAAAT	80274
Oy	1681	AAGAAAAAGAAACCATGTGAAGTCAAAATATTTGTTATTCAGTCAATTGGAATCTAT	1740
Db	80275	AAGAAAAAGAAACCATGTGAAGTCAAAATATTTGTTATTCAGTCAATTGGAATCTAT	80334
Oy	1741	TAAAAATATTTGAATCTTATATGATGAGAATCTCTTGACTCAATGAGACAGTGTGAG	1800
Db	80335	TAAAAATATTTGAATCTTATATGATGAGAATCTCTTGACTCAATGAGACAGTGTGAG	80394
Oy	1801	CTTTTTGGCGCTGTGGTCCCTACGTAAGAAAGAGGCTTTGTCAATAAGCTTATATGTAAC	1860
Db	80395	CTTTTTGGCGCTGTGGTCCCTACGTAAGAAAGAGGCTTTGTCAATAAGCTTATATGTAAC	80454
Oy	1861	AGGTGCCAAGTAATGATGCCAAGCTTGCTCTTAAAGCAATACGTAATTTGTTTATGACT	1920
Db	80455	AGGTGCCAAGTAATGATGCCAAGCTTGCTCTTAAAGCAATACGTAATTTGTTTATGACT	80514
Oy	1921	TTTATGTAAGTGAAGGGAATTAACAATCCCTCTGGAGAACTTCTCTCAATCCTTGGT	1980
Db	80515	TTTATGTAAGTGAAGGGAATTAACAATCCCTCTGGAGAACTTCTCTCAATCCTTGGT	80574
Oy	1981	GAACTCAATCTGCCAAGATTC	2001
Db	80575	GAACTCAATCTGCCAAGATTC	80595

RESULT 2  
AAV83942  
ID AAV83942 standard; DNA; 18443 BP.

```

XX AC AAV83942;
XX XX
XX DT 03-MAR-1999 (first entry)
XX XX
XX DE Bacterial artificial chromosome (BAC)-F2 contig 2.
XX XX
XX KW Yeast artificial chromosome; YAC; probe: eukaryotic chromosome;
XX KW neocentromere; replication; extra-chromosomal element; segregation;
XX KW cell division; artificial chromosome; gene therapy; BAC; transgenic;
XX KW human artificial chromosome; bacterial artificial chromosome; ss.
XX OS Synthetic.
XX FN W09851790-A1.
XX PD 19-NOV-1998.
XX PE 13-MAY-1998; 98WO-AU000352.
XX XX
XX PR 13-MAY-1997; 97AU-00006784.
XX PR 26-AUG-1997; 97AU-00008791.
XX PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX PI Choo K, Du Sart D, Cancellia MR;
XX DR WP1; 1999-009773/01.
XX XX
XX PT New isolated nucleic acid comprising neocentromere sequences from
XX PT eukaryotic chromosome - used to produce replicable, segregating
XX PT artificial chromosomes that can carry large amounts of DNA for gene
XX PT therapy.
XX PS
XX Claim 10; Page 181-194; 540pp; English.
XX
CC The present sequence represents a bacterial artificial chromosome (BAC)
CC contig, and exemplifies the invention. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in a
CC compatible cell, of replicating, acting as extra-chromosomal element and
CC segregating during cell division. The sequences can be used to construct
CC artificial chromosomes for use in gene therapy comprising a replicable,
CC segregating nucleic acid that confers a specific phenotype on cells.
CC Human artificial chromosomes can propagate in human cells and carry large
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC they are not mutagenic. The artificial chromosomes are also useful for
CC generation of transgenic plants and animals, in production of proteins
CC and to make diagnostic reagents, e.g. for expression of cytokines,
CC receptors and growth factors, or to increase the copy number of a gene in
CC a cell. The constructs may also be used for functional and structural
CC analysis of chromosomes
XX
SQ Sequence 18443 BP; 5380 A; 3651 C; 3772 G; 5627 T; 0 U; 13 Other;
OY Query Match 99.2%; Score 1984.8; DB 2; Length 18443;
DB Best Local Similarity 99.8%; Pred. No. 0;
OY 1 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGCTGAAAAATATATTTAACGTGCATTAAGT 60
DB 4292 ATCTTTATTTGTATGAGAAAAAGAAAAAGAGAGTAAATATTAACGTGATTTGT 4351
OY 61 TCAGAG-CGATGAGTTGCAAGTGACAGAAACTCAATCAAAACCAAGTAAGTCAAAAGGA 119
DB 4352 TCAGAGACCTTGAGATTGCAAGTAGACAGAAACTCAATCAAAACCAAGTAAGTCAAAAGGA 4411
OY 120 AATATATATGGCTCATGTAAACCTTCTCAGACAGAGGGCAGAGATGAGAAAGGGCTTTGGGAA 179
DB 4412 AATATATATGGCTCATGTAAACCTTCTCAGACAGAGGGCAGAGATGAGAAAGGGCTTTGGGAA 4471
OY 180 CAAGGAATTGTTCTCAAAATTCTAGAAATCTAGAGATTAGTCCAGAGATGGGTCAACCTTCC 239

```

4472 CAAGAAATGTTCTCAAAATCTAGAAATATAGATTAGTCCAGAGTGGTCACTTC 4531  
240 TGTCCCTGAGTGTGTGTGCGCATGGTAGAGTCTTATGGAGAGAAAGATGTTAG 299  
4532 TGTCCCTGAGTGTGTGTGCGCATGGTAGAGTCTTATGGAGAGAAAGATGTTAG 4591  
300 ATGAAGTAGGGCTAGCAAAACAAGGCAAGGGCACTATATCATCTAAAAATGGTTT 359  
4592 ATGAAGTAGGGCTAGCAAAACAAGGCAAGGGCACTATATCATCTAAAAATGGTTT 4651  
360 TTTTATGTCTCTCTTAATTTCACAATGCTTCCAAAGTAGACACACAGAAAAAGAA 419  
4652 TTTTATGTCTCTCTTAATTTCACAATGCTTCCAAAGTAGACACACAGAAAAAGAA 4711  
420 CATAGGACCTATACGTGGGAGCTTTTATCTTAAGCTTGTACTTGTCTTACAGCTT 479  
4712 CATAGGACCTATACGTGGGAGCTTTTATCTTAAGCTTGTACTTGTCTTACAGCTT 4771  
480 ACTCATGTCTTACCTGAGGCCATATGCCCTGTAAAGCTTGTGAGGGTTTCTATTA 539  
4772 ACTCATGTCTTACCTGAGGCCATATGCCCTGTAAAGCTTGTGAGGGTTTCTATTA 4831  
540 GCTGGGTTCTTATATAGGCTCTCTCCATTTCTGTGCTCATCTATAGATCTTTCTCT 599  
4832 GCTGGGTTCTTATATAGGCTCTCTCCATTTCTGTGCTCATCTATAGATCTTTCTCT 4891  
600 TTTTCTCACCTCTAGGACGTGGGCTGTTGTATGACGTGCTTAGGCTTGTGGGTT 659  
4892 TTTTCTCACCTCTAGGACGTGGGCTGTTGTATGACGTGCTTAGGCTTGTGGGTT 4951  
660 TTTTCTGGGACAAATGTCTTCAGATTTATCTAGAACCAATAACTACACGCCA CTGGGC 719  
4952 TTTTCTGGGACAAATGTCTTCAGATTTATCTAGAACCAATAACTACACGCCA CTGGGC 5011  
720 AGGCTCTTCTCTCTCCAACTGACACATGTTCCAGGGCTTTCACTTATAGTTCAT 779  
5012 AGGCTCTTCTCTCTCCAACTGACACATGTTCCAGGGCTTTCACTTATAGTTCAT 5071  
780 GCATCTTGGCAAAAGAGGCTAGTTAACTATAGCAATCTTAGCAATGATTTCTTTT 839  
5072 GCATCTTGGCAAAAGAGGCTAGTTAACTATAGCAATCTTAGCAATGATTTCTTTT 5131  
840 GACATGTTGTAGATCTATTCACATTTTGTATTAAGCAATTCCTCATAGAAACCAAC 899  
5132 GACATGTTGTAGATCTATTCACATTTTGTATTAAGCAATTCCTCATAGAAACCAAC 5191  
900 CGAAGCTAGAGTCTCTCTGGAATGAGGGTGGCTCTCTCAATACAGATGTTCTAGAGAG 959  
5192 CGAAGCTAGAGTCTCTCTGGAATGAGGGTGGCTCTCTCAATACAGATGTTCTAGAGAG 5251  
960 TGTATTTTGGGCACTTAATCTTCTCCACTTATAGGGGCAACACCTGAATTAACCA 1019  
5252 TGTATTTTGGGCACTTAATCTTCTCCACTTATAGGGGCAACACCTGAATTAACCA 5311  
1020 CTAAAGTTTGTCAATGTCATAGTTAGTCTCAGGACAGTGCAGCTCAGAGTGAACCTGA 1079  
5312 CTAAAGTTTGTCAATGTCATAGTTAGTCTCAGGACAGTGCAGCTCAGAGTGAACCTGA 5371  
1080 CCTTATGTGTGTCCAGCTTTCTTCTTCAAGAGTCAAGCTGTGTTTCTGCTGACTCT 1139  
5372 CCTTATGTGTGTGTCCAGCTTTCTTCTTCAAGAGTCAAGCTGTGTTTCTGCTGACTCT 5431  
1140 CCAATAGGAACATAGCTGGAATCCCAAGCCCAATCTGGAATAGTAAGTGTCTCTGAC 1199  
5432 CCAATAGGAACATAGCTGGAATCCCAAGCCCAATCTGGAATAGTAAGTGTCTCTGAC 5491  
1200 AGTCTAGAAAGTGTCTACCGCTGATCTCCAAAGCTGTGACACACCTGAGAGAGAAA 1259  
5492 AGTCTAGAAAGTGTCTACCGCTGATCTCCAAAGCTGTGACACACCTGAGAGAGAAA 5551  
1260 TGAAGAAAGTGGGCTCTTCAAGTAATCTTGTCTTTTCAACAAGCCCTTAATTTTACTGC 1319  
5552 TGAAGAAAGTGGGCTCTTCAAGTAATCTTGTCTTTTCAACAAGCCCTTAATTTTACTGC 5611

1320 ATAAATTTTGAATTCATGATTAATTTCTACAATTTTCCATTAAGTCACTTACACAA 1379  
5612 ATAAATTTTGAATTCATGATTAATTTCTACAATTTTCCATTAAGTCACTTACACAA 5671  
1380 TACCTCTCAGCAACACTTGGCTTGTCTTAATACATATCTATTTAGAGAGCTGTCTCT 1439  
5672 TACCTCTCAGCAACACTTGGCTTGTCTTAATACATATCTATTTAGAGAGCTGTCTCT 5731  
1440 TAAAGTAATGTTTATATAGCACTAAGGCTCTGGCTTACATATAAAGGGGTATTGAG 1499  
5732 TAAAGTAATGTTTATATAGCACTAAGGCTCTGGCTTACATATAAAGGGGTATTGAG 5791  
1500 CAATGTATACAGAAGCTTTTCTCCAGAGCTCTCATATGTAAAGAAATTCATTAGATTGG 1559  
5792 CAATGTATACAGAAGCTTTTCTCCAGAGCTCTCATATGTAAAGAAATTCATTAGATTGG 5851  
1560 CTGAATATAGACTGATCTGTCATTTTCTGCTCACTTATCATATAGGAAGTCAATTAGCTTA 1619  
5852 CTGAATATAGACTGATCTGTCATTTTCTGCTCACTTATCATATAGGAAGTCAATTAGCTTA 5911  
1620 GGAACAAAACCTACATCTATGTATTTAGAAACAAGCTGGTTTGTCAATATTAATA 1679  
5912 GGAACAAAACCTACATCTATGTATTTAGAAACAAGCTGGTTTGTCAATATTAATA 5971  
1680 TAAAGAAAAGAAACCATGTGAAGTCAAAATATTTGTTAATCAGTCAATTGAGATCTTA 1739  
5972 TAAAGAAAAGAAACCATGTGAAGTCAAAATATTTGTTAATCAGTCAATTGAGATCTTA 6031  
1740 TTAAGAAATGATTTGTAATCTTTATATAGAGAACTATCTTGACTCAAGTGACAGTGCTCA 1799  
6032 TTAAGAAATGATTTGTAATCTTTATATAGAGAACTATCTTGACTCAAGTGACAGTGCTCA 6091  
1800 GCTTTTGGCCGTGTGTCCTCATAGTGAAGGAGGCTTGTCAATAGCTTATATAGTA 1859  
6092 GCTTTTGGCCGTGTGTCCTCATAGTGAAGGAGGCTTGTCAATAGCTTATATAGTA 6151  
1860 CAGTGCCCAAGTTAAGTGCCCAAGCTTGTCTTAAAGCACTAGATTTTGTATTAGAC 1919  
6152 CAGTGCCCAAGTTAAGTGCCCAAGCTTGTCTTAAAGCACTAGATTTTGTATTAGAC 6211  
1920 TTTTATGTAAGTGAAGGAAATTAACAATTCCTCTGGAGAACTTCTCTCATCTTGG 1979  
6212 TTTTATGTAAGTGAAGGAAATTAACAATTCCTCTGGAGAACTTCTCTCATCTTGG 6271  
1980 TGAAGCATTTCTGCCAAGATTC 2001  
6272 TGAAGCATTTCTGCCAAGATTC 6293

RESULT 3  
AAV83940  
ID AAV83940 standard; DNA; 80240 BP.  
XX  
XX AAV83940;  
AC  
XX  
XX  
DT 03-MAR-1999 (first entry)  
XX  
DE NC-config derived from mardel(10) on chromosome 10q25.2.  
XX  
XX Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;  
XX neocentromere; replication; extra-chromosomal element; segregation;  
XX cell division; artificial chromosome; gene therapy; mardel(10);  
XX human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.  
OS Homo sapiens.  
XX  
XX  
XX PN M09851790-A1.  
XX  
XX  
PD 19-NOV-1998.  
XX  
XX 13-MAY-1998; 98WO-AU000352.  
XX  
XX

PR 13-MAY-1997; 97AU-00006784.  
PR 26-AUG-1997; 97AU-00008791.  
XX  
XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
PA  
PI Choo K, Du Sart D, Cancilla MR;  
XX  
XX WPI; 1999-009773/01.  
DR  
XX  
XX New isolated nucleic acid comprising neocentromere sequences from  
PT eukaryotic chromosome - used to produce replicable, segregating  
PT artificial chromosomes that can carry large amounts of DNA for gene  
PT therapy.  
XX  
XX  
PS Claim 9; Fig 16A; 540pp; English.  
XX  
XX The present sequence represents the NC-contig derived from a mutated  
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual  
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is  
CC mitotically stable and contains a functional neocentromere at a location  
CC regarded as non-centromeric. This neocentromere maps to q25.2 on  
CC chromosome 10. The specification describes nucleic acid sequences derived  
CC from a eukaryotic chromosome, including a neocentromere or its functional  
CC derivative or hybrid, that are able, in a compatible cell, of  
CC replicating, acting as extra-chromosomal element and segregating during  
CC cell division. The sequences can be used to construct artificial  
CC chromosomes for use in gene therapy comprising a replicable, segregating  
CC nucleic acid that confers a specific phenotype on cells. Human artificial  
CC chromosomes can propagate in human cells and carry large amounts of DNA  
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not  
CC transgenic. The artificial chromosomes are also useful for generation of  
CC transgenic plants and animals, in production of proteins and to make  
CC diagnostic reagents, e.g. for expression of cytokines, receptors and  
CC growth factors, or to increase the copy number of a gene in a cell. The  
CC constructs may also be used for functional and structural analysis of  
CC chromosomes  
XX  
XX  
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;  
Query Match 94.7%; Score 1895.2; DB 2; Length 80240;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1908; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
QY 1 ATCTTATTTGTGTAAGAAAAGAGAGTAAATATATTAACCTGATATAGT 60  
DB 78329 ATCTTATTTGTGTAAGAAAAGAGAGTAAATATATTAACCTGATATAGT 78388  
QY 61 TCAGA-CCATGATTGCAAGTACAGAACTCAATCAACCAAGTAAAGTCAAAAGA 119  
DB 78389 TCAGAACCTTGATGCAAGTACAGAACTCAATCAACCAAGTAAAGTCAAAAGA 78448  
QY 120 AAATATATTTGGCTCATGTAACTTTCTCAGAGAGGGCAGATGAAAGGGCTTTGGGA 179  
DB 78449 AAATATATTTGGCTCATGTAACTTTCTCAGAGAGGGCAGATGAAAGGGCTTTGGGA 78508  
QY 180 CAAGAGATTTGTTCTCAAAATTTCTAGGAATCTAGGATTAAGTCAAGATGGCTCACTTCC 239  
DB 78509 CAAGAGATTTGTTCTCAAAATTTCTAGGAATCTAGGATTAAGTCAAGATGGCTCACTTCC 78568  
QY 240 TGTCCCTGAGGTGGTGGTAGAGTGGTGAAGTCTTATGAGGAAAGAGTCAATGTTAG 299  
DB 78569 TGTCCCTGAGGTGGTGGTAGAGTGGTGAAGTCTTATGAGGAAAGAGTCAATGTTAG 78628  
QY 300 ATGAAGTGAAGGCTTAAGCAAAAGAGGCAAGGCCCATATATCATGTCTAAATAATGGTTT 359  
DB 78629 ATGAAGTGAAGGCTTAAGCAAAAGAGGCAAGGCCCATATATCATGTCTAAATAATGGTTT 78688  
QY 360 TTTTGAATGCTTCTTAATTTCAAAATGCTTCCAAAGTGAAGTCAAGCAAGCAAGAAAGAA 419  
DB 78689 TTTTGAATGCTTCTTAATTTCAAAATGCTTCCAAAGTGAAGTCAAGCAAGCAAGAAAGAA 78748  
QY 420 CATAGGACTCTAGTGGGTCTTTTATCTTAAGCCTGTGATCTGCTTTTCAAGCTT 479  
|||||

DB 78749 CATAGGACTCTAGTGGGTCTTTTATCTTAAGCCTGTGATCTGCTTTTCAAGCTT 78808  
QY 480 ACTACGCTTGATACCTGAGGCCATATGCCCTGTAAAGCTTGTGAGGGTTTCTACTA 539  
DB 78809 ACTACGCTTGATACCTGAGGCCATATGCCCTGTAAAGCTTGTGAGGGTTTCTACTA 78868  
QY 540 GCTGGGTTCTTATATGAGCTCTCTCCCATTTTCTGTTGCCCTCACTAGTATCTTCTCT 599  
DB 78869 GCTGGGTTCTTATATGAGCTCTCTCCCATTTTCTGTTGCCCTCACTAGTATCTTCTCT 78928  
QY 600 TTTTCTACCTCTGAGGACTGGTGGCTGTTTGTATGAGACTGCTTACCTTGTGCTTGGGTT 659  
DB 78929 TTTTCTACCTCTGAGGACTGGTGGCTGTTTGTATGAGACTGCTTACCTTGTGCTTGGGTT 78988  
QY 660 TTTTCTGGGAGCAATGCTCTTCAAGATTAATCTGACCAATTAACCTACAGCCCTGGGCC 719  
DB 78989 TTTTCTGGGAGCAATGCTCTTCAAGATTAATCTGACCAATTAACCTACAGCCCTGGGCC 79048  
QY 720 AGGCTCTGCTCTCCCAAGCTGAGCAATGTTCCAGGGCTCTTACCTTATGTTAGGTCAA 779  
DB 79049 AGGCTCTGCTCTCCCAAGCTGAGCAATGTTCCAGGGCTCTTACCTTATGTTAGGTCAA 79108  
QY 780 GCATCTTGAGCAAAAGAAAGGCTTAAATTAAGCAATTTAGCAATTTGATCTTTT 839  
DB 79109 GCATCTTGAGCAAAAGAAAGGCTTAAATTAAGCAATTTAGCAATTTGATCTTTT 79168  
QY 840 GACATGTTGATATCTATTCATTTTGTATTAAGCATTCCTTATGAAACCAACA 889  
DB 79169 GACATGTTGATATCTATTCATTTTGTATTAAGCATTCCTTATGAAACCAACA 79228  
QY 900 CGAATACGCTGCTCGAAGAGGAGGAGGCTCTCAATCAGAGTGTCTAGAGAG 959  
DB 79229 CGAATACGCTGCTCGAAGAGGAGGAGGCTCTCAATCAGAGTGTCTAGAGAG 79288  
QY 960 TGTATTTTGGGCACTTAATCTTCTCACTACTTAAAGGCAAGCACTGAATTAACCA 1019  
DB 79289 TGTATTTTGGGCACTTAATCTTCTCACTACTTAAAGGCAAGCACTGAATTAACCA 79348  
QY 1020 CTAAATTTGTCATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079  
DB 79349 CTAAATTTGTCATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 79408  
QY 1080 CCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1139  
DB 79409 CCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 79468  
QY 1140 CCATAGGAACATCAGTCTGAAATCTCAGACCAACATCTGAGTAAAGTCTCTGAC 1199  
DB 79469 CCATAGGAACATCAGTCTGAAATCTCAGACCAACATCTGAGTAAAGTCTCTGAC 79528  
QY 1200 AGTCTTGAAGTGTGCTTACCGCTGATCTTCMAAGCTGTGACACACCGTGAGAGAA 1259  
DB 79529 AGTCTTGAAGTGTGCTTACCGCTGATCTTCMAAGCTGTGACACACCGTGAGAGAA 79588  
QY 1260 TGAAGAAAGCTGGCTCTTCAAGTAAATCTTGTCTTTTCAAGGCCCTTAATTTTCTG 1319  
DB 79589 TGAAGAAAGCTGGCTCTTCAAGTAAATCTTGTCTTTTCAAGGCCCTTAATTTTCTG 79648  
QY 1320 ATATATTTTGAATTCATGATTAATTTCTACAATTTTCCATTAAGTCACTACACAA 1379  
DB 79649 ATATATTTTGAATTCATGATTAATTTCTACAATTTTCCATTAAGTCACTACACAA 79708  
QY 1380 TACCTCTCATGCAACATTTGCTTGTATTAATCATATCTATTTAGAGCTGTGCTTCT 1439  
DB 79709 TACCTCTCATGCAACATTTGCTTGTATTAATCATATCTATTTAGAGCTGTGCTTCT 79768  
QY 1440 TAAAGCTAATGTTTATATGACTAAGGCTCTTGGCTTAACATTAAGGGGATTTGAG 1499  
DB 79769 TAAAGCTAATGTTTATATGACTAAGGCTCTTGGCTTAACATTAAGGGGATTTGAG 79828  
QY 1500 CAATGATATACAGAACTTTTCTCCAGAGGCTCATATGATTAAGAAATTCATTAAGTTGG 1559  
DB 79829 CAATGATATACAGAACTTTTCTCCAGAGGCTCATATGATTAAGAAATTCATTAAGTTGG 79888  
|||||



QY 1560 CTGAATATAGCTGATCTGTCATTTCTGCTCATTTATCATAGAGATTTAGCTAA 1619  
| | | | |  
DB 79889 CTGAATATAGCTGATCTGTCATTTCTGCTCATTTATCATAGAGATTTAGCTAA 79948  
| | | | |  
QY 1620 GGAAACAAAACCTACATCTATGTATATAGAGAACAAAGCTGTTTGTCTCATATATAAA 1679  
| | | | |  
DB 79949 GGAAACAAAACCTACATCTATGTATATAGAGAACAAAGCTGTTTGTCTCATATATAAA 80008  
| | | | |  
QY 1680 TAAGAAAAAGAACCCATGTGAAGTCAAAATATTGTTTATAGGTCATTTGATTTCTTA 1739  
| | | | |  
DB 80009 TAAGAAAAAGAACCCATGTGAAGTCAAAATATTGTTTATAGGTCATTTGATTTCTTA 80068  
| | | | |  
QY 1740 TTAATAAGATTTGAATTTCTTTATGATGAGAACTATCTGACATCAAGTGACAGTGTGA 1799  
| | | | |  
DB 80069 TTAATAAGATTTGAATTTCTTTATGATGAGAACTATCTGACATCAAGTGACAGTGTGA 80128  
| | | | |  
QY 1800 GCTTTTGGCTGTGTGCTCCCTACGTAGAAAGAGCGTTTGTCTATTAAGCTTTATGTGA 1859  
| | | | |  
DB 80129 GCTTTTGGCTGTGTGCTCCCTACGTAGAAAGAGCGTTTGTCTATTAAGCTTTATGTGA 80188  
| | | | |  
QY 1860 CAGGTGCCAAGTTAAGTCCCAAGCTTGTCTTAAGACATACCTGATTTTG 1911  
| | | | |  
DB 80189 CAGGTGCCAAGTTAAGTCCCAAGCTTGTCTTAAGACATACCTGATTTTG 80240  
| | | | |  
RESULT 4  
AAV83958  
ID AAV83958 standard; DNA; 300 BP.  
XX  
AC AAV83958;  
XX  
DT 03-MAR-1999 (first entry)  
XX  
DE Bacterial artificial chromosome (BAC)-F2 contig 47 fragment 2.  
XX  
XX Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;  
XX neocentromere; replication; extra-chromosomal element; segregation;  
XX cell division; artificial chromosome; gene therapy; BAC; transgenic;  
XX human artificial chromosome; bacterial artificial chromosome; ss.  
XX  
OS Synthetic.  
XX  
PN WO9851790-A1.  
XX  
PD 19-NOV-1998.  
XX  
XX 13-MAY-1998; 98WO-AU000352.  
XX  
XX 13-MAY-1997; 97AU-00006784.  
XX  
PR 26-AUG-1997; 97AU-00008791.  
XX  
PA (AMRA-) AMRAD OPERATIONS PTY LTD.  
XX  
PI Choo K, Du Sart D, Cancilla MR;  
XX  
DR WPI; 1999-009773/01.  
XX  
PT New isolated nucleic acid comprising neocentromere sequences from  
PT eukaryotic chromosome - used to produce replicable, segregating  
PT artificial chromosomes that can carry large amounts of DNA for gene  
PT therapy.  
XX  
PS Claim 10; Page 235; 540pp; English.  
XX  
CC The present sequence represents a bacterial artificial chromosome (BAC)  
CC contig, and exemplifies the invention. The specific contig describes  
CC nucleic acid sequences derived from a eukaryotic chromosome, including a  
CC neocentromere or its functional derivative or hybrid, that are able, in a  
CC compatible cell, of replicating, acting as extra-chromosomal element and  
CC segregating during cell division. The sequences can be used to construct  
CC artificial chromosomes for use in gene therapy comprising a replicable,  
CC segregating nucleic acid that confers a specific phenotype on cells.

CC Human artificial chromosomes can propagate in human cells and carry large  
CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,  
CC they are not mutagenic. The artificial chromosomes are also useful for  
CC generation of transgenic plants and animals, in production of proteins  
CC and to make diagnostic reagents, e.g. for expression of cytokines,  
CC receptors and growth factors, or to increase the copy number of a gene in  
CC a cell. The constructs may also be used for functional and structural  
CC analysis of chromosomes  
XX  
SQ Sequence 300 BP; 54 A; 70 C; 56 G; 101 T; 0 U; 19 Other;  
Query Match 4.9%; Score 97.6; DB 2; Length 300;  
Best Local Similarity 66.7%; Pred. No. 3.4e-15;  
Matches 168; Conservative 18; Mismatches 62; Indels 4; Gaps 4;  
QY 1753 GAATCTTATAGAGAACTATCTGACTCAAGTGACAGTGTGAGCTTTTGAGCCG 1812  
| | | | |  
DB 1 GAATCTCTGAAATTAACAATCTATCTTGTCTCAAAATTSACTGTAGCTAACTGGCCG 60  
| | | | |  
QY 1813 TGGTCCCTACG-TAGAAAGAGGCTTTGTATTAAG-TCTTATATGTACAGTCCCAAG 1870  
| | | | |  
DB 61 TGGTCCCTTGGCTTTAATGAGGCTTTGTATATATGATCATATGATGTGCTGCTGCTAG 120  
| | | | |  
QY 1871 TT-AGTGGCCCAAGTGTCTCTTAAGACATCTGATTTTGTTAAGCTTTTATGTGA 1929  
| | | | |  
DB 121 TTGTAGTGCCCTGCTGTCTTGTCTGCTTACTGATTTWGGGTATACATCATATATTA 180  
| | | | |  
QY 1930 CTGAAGGGAATTAACAATCTCTGAGGAGAACTTCTCCATCTTGTGTAAGTCAAT 1989  
| | | | |  
DB 181 YTSMAAGTCT-TTCTCTCTCCGGYGGAGAAATTTCTCTCTCTCGGAGAACTCTT 239  
| | | | |  
QY 1990 CTGCCGAATTC 2001  
| | | | |  
DB 240 CTSCGGAATTC 251  
| | | | |  
RESULT 5  
ADA71938  
ID ADA71938 standard; DNA; 2000 BP.  
XX  
AC ADA71938;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Rice gene, SEQ ID 5263.  
XX  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
XX gene; db.  
XX  
OS Oryza sativa.  
XX  
PN MO2003000898-A1.  
XX  
PD 03-JAN-2003.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
XX  
PR 22-JUN-2001; 2001WO-IB001105.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX  
DR WPI; 2003-175290/17.  
XX  
PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
PS Claim 27; SEQ ID NO 5263; 899pp; English.  
XX



XX	Human secreted protein 5' EST, SEQ ID NO: 19284.
XX	Human; 5' EST, expressed sequence tag; secreted protein; cDNA isolation;
KW	gene therapy; chromosome mapping; ss.
XX	Homo sapiens.
XX	EP1033401-A2.
PN	06-SEP-2000.
PD	21-FEB-2000; 2000EP-00200610.
XX	26-FEB-1999; 99US-0122487P.
PR	(GEST ) GENSET.
XX	Dumas Milne Edwards J, Duclert A, Giordano J;
PI	WP1; 2000-500381/45.
XX	
XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX	
XX	Claim 1; SEQ ID NO 19284; 71pp + Sequence listing; English.
PS	
CC	The present sequence is one of a large number of 5' ESTs derived from
CC	mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC	identified within the present sequence. The 5' ESTs were prepared from
CC	total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC	sequences usually correspond mainly to the 3' untranslated region (UTR)
CC	of the mRNA because they are often obtained from oligo-dT primed cDNA
CC	libraries. Such ESTs are not well suited for isolating cDNA sequences
CC	derived from the 5' ends of mRNAs and even in those cases where longer
CC	cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC	ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC	to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC	diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC	They are used to obtain upstream regulatory sequences and to design
CC	expression and secretion vectors
XX	
XX	Sequence 431 BP; 109 A; 77 C; 112 G; 132 T; 0 U; 1 Other;
SO	
Query Match	2.5%; Score 49.4; DB 3; Length 431;
Best Local Similarity	71.4%; Pred. No. 0.02; Mismatches 26; Indels 0; Gaps 0;
Matches	65; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY	70 TGGATTGCAGAGTGCAGAAACTCAATTCAACCAACGTAAGTCAAAAGGAAATATATTG 129
DB	231 TTGGTTGAAGTGAAGACAAAGAAACCAACTCAATTAGCTTAGAGCCAAAAGAAATATATG 290
QY	130 GCTCATGTAACTTCTCAGACAGAGGCGCAG 160
DB	231 GTTCATGGAAACCAAACTAAGAAAAGGACG 321
RESULT 8	
ADZ70899/c	
ID	ADZ70899 standard; DNA; 1428 BP.
XX	
AC	ADZ70899;
XX	
XX	14-JUL-2005 (first entry)
DT	
XX	
DE	Human matrix attachment region DNA SEQ ID NO 23.
XX	
KW	ds; matrix attachment region; MAR; protein production.
XX	
OS	Homo sapiens.
XX	
PN	WO2005040377-A2.

[illegible]

## RESULT 9

ADZ70899 standard; DNA; 1428 BP.

ADZ70899;

14-JUL-2005 (first entry)

Human matrix attachment region DNA SEQ ID NO 23.

ds; matrix attachment region; MAR; protein production.

Homo sapiens.

WO2005040377-A2.

06-MAY-2005.

22-OCT-2004; 2004WO-EP011974.

24-OCT-2003; 2003US-0513574P.

06-FEB-2004; 2004EP-00002722.

(SELEX-) SELEXIS SA.

Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;

Puttini S;

WPI; 2005-333507/34.

New purified and isolated DNA sequence having protein production increasing activity comprises a bent DNA element and a binding site for a DNA binding protein, useful for increasing protein production activity in eukaryotic host cell.

Claim 4; SEQ ID NO 23; 282bp; English.

The invention relates to a purified and isolated DNA sequence having protein production increasing activity comprising at least one bent DNA element, and at least one binding site for a DNA binding protein. The purified and isolated DNA sequence comprising a first and a second isolated matrix attachment region (MAR) nucleotide sequence, which is a MAR nucleotide sequence selected from a purified and isolated DNA sequence above, a purified and isolated MAR DNA above, a purified and isolated clysmar element and/or fragment, a synthetic MAR sequence, a sequence complementary to it, its molecular chimera, or its combinations and variants, is useful for increasing protein production activity in a eukaryotic host cell. The present sequence represents a human matrix attachment region DNA.

Sequence 1428 BP; 652 A; 18 C; 6 G; 752 T; 0 U; 0 Other;

Query Match 2.2%; Score 45; DB 14; Length 1428;

Best Local Similarity 43.5%; Pred. No. 0.48; Mismatches 265; Indels 0; Gaps 0;

Matches 204; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

1310 ATTTTACTGCATATATATTTGAATTCACGTATATTTTCCATATGATC 1369

786 ATTATATATATATATATATATTTTATATATATATATATATATATTTTATTA 845

1370 CTACACCAATACCCTCTACGACACCTGGCTTGCTAATCATCTATTATGAG 1429

846 TAAATATATATATATATATATTTTATATATATATATATATATATATATAT 905

1430 CTGCTCTTAAAGCGTAATGTTTATATGACCTAAGCTTGGCTTACATATAA 1489

906 ATTTTATATATATATATATTTTATATATATATATATATATATATATATAT 965

1490 GGGTATGAGCAATGTCATACAGAGCTTTTCTCAGAGCTCATATGTAAGATTC 1549

966 ATATATATATATATATATATATATTTTATATATATATATATATATATATATA 1025

QY 1550 ATTAGATGGCTGAATAGATGATCGATCTGCTGCTCACTTATCATAGAAAGT 1609

Db 1026 TATATATTTTA 1085

QY 1610 CATTAGCTAAGAAACAAACCTACATCTATGTAATTTAGAAACAGCTGTTTGC 1669

Db 1086 TATTTTA 1145

QY 1670 AATATATAAATTAAGAAAGAAACCTGTCAGAAAGTCAAAATTTTGTATAGATC 1729

Db 1146 TA 1205

QY 1730 TGAGATCTATTAAGAAAGTATGTAATCTTATGATGAGACTATCTT 1778

Db 1206 ATTTT 1254

## RESULT 10

ABK28365 standard; DNA; 9238 BP.

ABK28365;

23-APR-2002 (first entry)

DNA transcription associated genomic DNA #120.

DNA transcription associated gene; peptide nucleic acid; PNA-oligomer; PNA; cytosine methylation state; SNP; retroviral infection; gene; ds; single nucleotide polymorphism; adenosine deaminase deficiency; cancer; viral infection; Sezary syndrome; haematological disorder; tuberculosis; immunological disorder; Werner syndrome; developmental disorder; psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis; neurodegenerative disorder; Maardenburg syndrome; Niemann-Pick disease; myelodysplastic syndrome; myocardial infarction; hypertension; arthritis; angiogenesis; congenital heart disease; HDR syndrome; gene therapy; polyglutamine disorder; solid tumour.

Unidentified.

WO200192565-A2.

06-DEC-2001.

06-APR-2001; 2001WO-EP003973.

06-APR-2000; 2000DE-01019058.

07-APR-2000; 2000DE-01019173.

30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-090046/12.

New nucleic acids or oligomers, useful for diagnosing or treating diseases associated with DNA transcription, e.g. immunological disorders, Werner syndrome, psoriasis, myocardial infarction, solid tumors or cancer.

Claim 1; SEQ ID NO 239; 32bp; English.

The invention relates to a nucleic acid, which comprises a segment of the chemically pretreated DNA of genes associated with DNA transcription from one of 346 sequences, and an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical to the chemically pretreated DNA of genes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for diagnosing or treating diseases associated with DNA transcription

CC (particularly with the methylation status), e.g. adenosine deaminase  
CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
CC haematological disorders, immunological disorders, Werner syndrome,  
CC tuberculousis, developmental disorders, psoriasis, Rieger's syndrome,  
CC neurodegenerative disorders, neurodegenerative disorders, Mardenburg  
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
CC associated genomic DNA molecules of the invention. Note: The sequence  
CC data for this patent did not form part of the printed specification but  
CC was obtained in electronic format directly from the European Patent  
CC Office

CC Sequence 9238 BP; 2692 A; 159 C; 1985 G; 4402 T; 0 U; 0 Other;

Query Match 2.2%; Score 44.8; DB 6; Length 9238;  
Best Local Similarity 56.9%; Pred. No. 1.2;  
Matches 82; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1611 ATTAGCTAAGGACAAACCTACATCTATGTAATTGAGAGCAAGCTGTTTGCCTCA 1670  
DB 1932 ATATTTTATTAAGAAAAGAAATATATGATATGAGAAATATGTAAGATCTTTATGTTA 1991  
QY 1671 ATATTAATAATAGAAAAGAACCATGTGAAGTCAAAATATTTGTTTAATCAAGTCATT 1730  
DB 1992 TTTTAAAAAATAGAAATATATTAATTAATTAATTTTATTTTATTTATTTGATT 2051  
QY 1731 GAGATCTATTAATAAGTATTGCA 1754  
DB 2052 GGTAAATTTTAAAAAATGTTGTA 2075

RESULT 11  
AAV20419/c  
ID AAV20419 standard; cDNA to mRNA; 3100 BP.

XX AAV20419;  
XX  
XX  
DT 15-JUN-1998 (first entry)  
XX  
DE Human discs large 1 gene-1 cDNA cancer related molecule.

XX Human; NE-dlg; discs large 1 gene; cancer related molecule; nerve;  
XX internal secretion tissue; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 234..2687  
FT /\*tag= a  
FT /product= "dlg 1"

XX JPI0066581-A.  
XX  
XX 10-MAR-1998.  
XX  
XX 23-AUG-1996; 96JP-00241370.  
XX  
XX 23-AUG-1996; 96JP-00241370.  
XX  
XX (SUME ) SUMITOMO ELECTRIC IND CO.  
XX  
XX WPI: 1998-224339/20.  
XX P-PSDB; AAM48101.

XX Human discs large 1 gene family - useful in, e.g. therapeutic  
PT composition(s) for treating cancer.

XX Claim 16; Page 17-18; 31pp; Japanese.

XX The present sequence encodes human dlg-1 molecule. The present invention  
CC describes human discs large 1 gene (dlg) family expressible in nerve

CC tissue. Also described are: (1) a polynucleotide (PN) encoding dlg and  
CC comprising a 3100 (I) or 3213 bp (II) sequence; (2) RNA corresponding to  
CC PN having sequences (I) and (II); (3) RNA having a base sequence  
CC translated by dlg; (4) an antisense PN having a at least 15 bp sequence,  
CC and which is a part of PN of (1); (5) a derivative of the antisense PN;  
CC (6) an antibody specific for dlg, and (7) an antibody specific for a  
CC polypeptide having a 817 or 849 aa sequence. The polypeptide, RNA s and  
CC antibodies can be used for detection of dlg. The antisense PN can be used  
CC as a therapeutic composition for treating cancer

CC Sequence 3100 BP; 753 A; 799 C; 902 G; 646 T; 0 U; 0 Other;

Query Match 2.2%; Score 44.4; DB 2; Length 3100;  
Best Local Similarity 48.4%; Pred. No. 0.96;  
Matches 123; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 1670 AATATTAATAATAGAAAAGAAACCATGTAAGTCAAAATATTTGTTTAACAGTCAAT 1729  
DB 2859 AACTTAAAAAAGAAAAGAAACGACTAGAAAACCTTAATTTTGTGACATATCTAAAAA 2800  
QY 1730 TGAGATCTATTTAAAAAGATTTGAATCTTTATGATGAGAACTATCTGACTCAAGTGG 1789  
DB 2799 AAGGGAGCAAGAACAGTAGATTTGTTTCCCATGGGGACAGAAATGAAGAGGAGGAGG 2740  
QY 1790 AAGTGTGAGCTTTTGGCCCTGTGCTCCCTACGTAGAAAGAGCTTTGTATTAAGTC 1849  
DB 2739 GAAGAGGAGCTTGATTTCTTCTGTTCCACAGAGAAATGTTGAGGGGATTTCTTCAGAGTT 2680  
QY 1850 TTAATAGTACAGGAGCCCAAGTTAAGTCCCAAGCTGCTTAAAGCATCTGAGATT 1909  
DB 2679 TTTCAAGGGATGGAGACCAATATGATGTCACAGACTGCTCTCAATGATTTGTTGATT 2620  
QY 1910 TGTTTTACACTTTT 1923  
DB 2619 TGTATTAATCTCT 2606

RESULT 12  
AAX86725/c  
ID AAX86725 standard; DNA; 3100 BP.

XX AAX86725;  
XX  
XX  
DT 27-OCT-1999 (first entry)  
XX  
DE DNA sequence encoding tumour suppressor protein NE-dlg.

XX Nedasin; tumour suppressor protein; NE-dlg; tumour suppression;  
XX malignant tumour formation; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 234..2687  
FT /\*tag= a

XX MO9943702-AI.  
XX  
XX 02-SEP-1999.  
XX  
XX 24-AUG-1998; 98WO-JP003740.  
XX  
XX 25-FEB-1998; 98JP-00043552.  
XX  
XX (SUME ) SUMITOMO ELECTRIC IND CO.

XX Kishimoto T, Niwa S, Kuwahara H, Saya H;  
PI WPI: 1999-51835/43.  
XX P-PSDB; AAY30137.

XX New nedasin protein, useful in the investigation of the mechanism of  
PT formation of malignant tumors.

REF ID: A66566

XX

XX





PF 22-JUN-2001; 2001WO-IB001105.  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,  
PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX  
XX WPI; 2003-175290/17.  
DR  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
XX Claim 27; SEQ ID NO 6508; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
XX Sequence 2000 BP; 715 A; 429 C; 393 G; 463 T; 0 U; 0 Other;  
SQ  
XX  
XX Query Match 2.2%; Score 44.2; DB 8; Length 2000;  
Best Local Similarity 56.6%; Pred. No. 0.9;  
Matches 82; Conservative 0; Mismatches 63; Indels 0; Gapd 0;  
QY 1608 GTGATTAGCTAAGACAACTACATCTATGTAATTAAGAGCAAGCTGTTTTC 1667  
DB 167 GTTAATATATTAATTCATAATTAATAATTAATTTGTAAGAGCTGAGTTCA 226  
QY 1668 TCAATATTAATAAATAGAAAAAGAACCATGCTGAAGTCAAAATATTTGTTTAACAGTTC 1727  
DB 227 TGTATATTAATAAATAATGCTTAATAATTTGAACAAAAAGAAATATTAATAAGATGG 286  
QY 1728 ATTGAGATCTATTAATAAGATTTT 1752  
DB 287 GTTTAGAGTCTTAAGAAATATAT 311  
XX  
XX RESULT 17  
ACLS35363  
ID ACLS35363 standard; cDNA; 2000 BP.  
XX  
XX ACLS35363;  
AC  
XX  
XX 02-JUN-2005 (first entry)  
DT  
XX  
XX Rice stress-regulated promoter SEQ ID NO:13926.  
DE  
XX  
XX ss; abiotic stress tolerance; transgenic plant; plant; cereal;  
XX agricultural.  
XX  
XX Oryza sativa.  
OS  
XX  
XX WO2003008540-A2.  
PN  
XX  
XX 30-JAN-2003.  
PD  
XX  
XX 21-JUN-2002; 2002WO-US019668.  
PF  
XX  
XX 22-JUN-2001; 2001US-0300112P.  
PR  
XX  
XX 24-AUG-2001; 2001US-0314662P.  
PR  
XX  
XX 26-SEP-2001; 2001US-0325277P.  
PR  
XX  
XX 21-NOV-2001; 2001US-0332132P.  
PR

XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA  
XX  
XX Krepes J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F,  
PI Moughamer T, Provart N, Rieke D, Zhu T;  
XX  
XX WPI; 2003-248011/24.  
DR  
XX  
XX New stress-responsive nucleic acid, useful for altering the  
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold  
PT stresses, salt stress or osmotic stress.  
XX  
XX Claim 48; SEQ ID NO 13926; 89pp; English.  
XX  
XX The invention relates to novel abiotic stress responsive polynucleotides  
CC and polypeptides. Also disclosed are vectors, expression cassettes, host  
CC cells, and plants containing such polynucleotides. Also disclosed are  
CC methods for using the polynucleotides and polypeptides to alter the  
CC responsiveness of a plant to abiotic stress. The invention is useful in  
CC agriculture. The nucleic acid is useful for determining whether a test  
CC plant has been exposed to an abiotic stress condition. It is also useful  
CC for selecting an agent that alters abiotic stress regulated  
CC polynucleotide expression in a plant cell, and to identify a homolog or  
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid  
CC molecule and the polypeptide encoded by it are useful in altering the  
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt  
CC stress, osmotic stress or any of their combinations. The present sequence  
CC is used in the exemplification of the invention.  
XX  
XX Sequence 2000 BP; 540 A; 344 C; 416 G; 428 T; 0 U; 272 Other;  
SQ  
XX  
XX Query Match 2.2%; Score 44.2; DB 11; Length 2000;  
Best Local Similarity 13.0%; Pred. No. 0.9;  
Matches 37; Conservative 130; Mismatches 118; Indels 0; Gaps 0;  
QY 1236 GTGTGACACACCGTGTGAGAAAGAAAGTGGGCTCTTCAAGTAATCTTGGTTT 1295  
DB 36 GAKWMTMSAYMYAMMYAMMYAKKKMTTRAKCTSSKMSASAKRMMWMSMARMSKM 95  
QY 1296 TCACAAAGCCCTTAATTTTATGTCATTAATTAATTTGTAATTAATTTCTACAAAT 1355  
DB 96 TYMGGRMRYKAYMYRMWYMMWMAWMSWTRMYMNACTYCTMYMYWYMAAAMRW 155  
QY 1356 TTCCCATAGTATCTACACAAATACCTCTCATGCAACACTGTGCTTGTCAATACAT 1415  
DB 156 KMYKMMWARGWSAKMYAYRRKMYCTYMMARMAAMMTTKMCTMSWYMMTCYATWMAW 215  
QY 1416 ATCTATTTATGAGAGCTGCTCTTAAAGCGTAATGTTTATATGACACTAAGGCTCTTG 1475  
DB 216 ATATWTTTTRMAAMWKTTRAMTMYRRYAMMYTMMWAKKMAKATKTTTYAMRMMWMSMR 275  
QY 1476 CTTACATATPAAAGGGGTATGAGCAATGTGATACAGAACTTT 1520  
DB 276 TTYRMWMTWMMYWTWSWYTTAAMSWMYWCMTWSRRKMYRYK 320  
XX  
XX RESULT 18  
ABV72623/c  
ID ABV72623 standard; DNA; 82938 BP.  
XX  
XX ABV72623;  
AC  
XX  
XX 27-NOV-2002 (first entry)  
DT  
XX  
XX Human transporter protein encoding gene.  
DE  
XX  
XX Human; transporter protein; htp; gamma-aminobutyric-acid; GABA;  
XX gene therapy; protein therapy; vaccine; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH  
XX  
XX CDS 1120. .81932  
FT

FT	/*tag= a		FT	/*tag= x
FT	/product= "Transporter protein"		FT	/standard_name= "Single nucleotide polymorphism"
FT	/note= "Contains 8 introns"		FT	replace(19307. .19308,TMC)
FT	1120. .1223		FT	/*tag= y
FT	/*tag= b		FT	/standard_name= "Single nucleotide polymorphism"
FT	/number= 1		FT	replace(19308. .19309,CCC)
FT	1224. .27733		FT	/*tag= z
FT	/*tag= c		FT	/standard_name= "Single nucleotide polymorphism"
FT	/number= 1		FT	replace(19313. .19314,TWA)
FT	replace(2419, C)		FT	/*tag= aa
FT	/*tag= d		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(19314. .19315,AWT)
FT	replace(3191,A)		FT	/*tag= ab
FT	/*tag= e		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(19315. .19316,TCC)
FT	replace(4195,T)		FT	/*tag= ac
FT	/*tag= f		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(19316. .19317,CYT)
FT	replace(5910,C)		FT	/*tag= ad
FT	/*tag= g		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(19317. .19318,TWA)
FT	replace(7119,T)		FT	/*tag= ae
FT	/*tag= h		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(19318. .19319,ATT)
FT	replace(7529,G)		FT	/*tag= af
FT	/*tag= i		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(19860,G)
FT	replace(8855,C)		FT	/*tag= ag
FT	/*tag= j		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(20843,G)
FT	replace(8956,A)		FT	/*tag= ah
FT	/*tag= k		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(21144,T)
FT	replace(11628,C)		FT	/*tag= ai
FT	/*tag= l		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(21371,G)
FT	replace(11670,A)		FT	/*tag= aj
FT	/*tag= m		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(22470. .22471,A)
FT	replace(11671,G)		FT	/*tag= ak
FT	/*tag= n		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(22472. .22473,A)
FT	replace(11931,C)		FT	/*tag= al
FT	/*tag= o		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(22321,T)
FT	replace(15984,G)		FT	/*tag= am
FT	/*tag= p		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(225284,A)
FT	replace(17735,A)		FT	/*tag= an
FT	/*tag= q		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(25284,A)
FT	replace(17910,A)		FT	/*tag= ao
FT	/*tag= r		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(25934,G)
FT	replace(18130,C)		FT	/*tag= ap
FT	/*tag= s		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(27509. .27510,TAT)
FT	replace(19293. .19294,TWA)		FT	/*tag= ar
FT	/*tag= t		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(27513. .27514,TTT)
FT	replace(19294. .19295,ATT)		FT	/*tag= as
FT	/*tag= u		FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name=		FT	replace(27734. .27735,ATC)
FT	replace(19296. .19297,CYT)		FT	/*tag= at
FT	/*tag= v		FT	/number= 2
FT	/standard_name=		FT	27883. .41045
FT	replace(19303. .19304,ATC)		FT	/*tag= au
FT	/*tag= w		FT	/number= 2
FT	/standard_name=		FT	replace(28608,G)
FT	replace(19304. .19305,CYC)		FT	/*tag= av

```
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28627,C)
FT /*tag= aw
FT /standard_name= "Single nucleotide polymorphism"
FT replace(29709,T)
FT /*tag= ax
FT /standard_name= "Single nucleotide polymorphism"
FT replace(25788,T)
FT /*tag= ay
FT /standard_name= "Single nucleotide polymorphism"
FT replace(29878,T)
FT /*tag= az
FT /standard_name= "Single nucleotide polymorphism"
FT replace(30294,T)
FT /*tag= ba
FT /standard_name= "Single nucleotide polymorphism"
FT replace(30896,.30897,T)
FT /*tag= bb
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31815,G)
FT /*tag= bc
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31999,TTT)
FT /*tag= bd
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32026,T)
FT /*tag= be
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32146,A)
FT /*tag= bf
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32184,S)
FT /*tag= bg
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32214,.32215,T)
FT /*tag= bh
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32215,C)
FT /*tag= bi

Query Match
Best Local Similarity 5.2%; Score 44.2; DB 6; Length 82938;
Matches 85; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1623 ACAAACATCAATCTATGTAATTAAGAAACAAGCTGTTGCTCAATATAAATAA 1682
    |||||
DB 17088 ACAAACAGGAGGATTAATCAATAGGTTCTTAGACATTCGACATTTTAAATATTT 17029
    |||||

QY 1683 GAAAAAGAACCATGTGAAGTCAAAATATTTGTTAATGAGTCAATGGAATCTATTA 1742
    |||||
DB 17028 GTCAGAAAGAACTTGAATAATACAGAAAAATTAATTAATTAATGAATAATCTTGG 16969
    |||||

QY 1743 AAAAGTATTTGAATCTTTATGATGAGAACTAT 1775
    |||||
DB 16968 GAAATATCTCTGTTAAATGATGTAAATAGTAT 16936
    |||||

RESULT 19
ABX13938
ID ABX13938 standard; DNA; 3891 BP.
XX
AC ABX13938;
XX
DE 03-MAR-2003 (first entry)
XX
CLOSTRIDIUM BOTULINUM, botulinum neurotoxin serotype A DNA.
XX
KW Botulinum toxin; serotype A; neurotoxin; needleless syringe;
    involuntary muscle contraction; spasmodic dysphonia; laryngeal dystonia;
    oromandibular dysphonia; lingual dystonia; cervical dystonia;
    focal hand dystonia; blepharospasm; strabismus; hemifacial spasm;
    eyelid disorder; cerebral palsy; focal spasticity; spasmodic colitis;
    neurogenic bladder; anismus; limb spasticity; tic; tremor; bruxism;
```

```
KM anal fissure; achalasia; fibromyalgia; dysphagia; wrinkle; brow furrow;
KW vaccine; motor function; cross-contamination; blood-borne pathogen;
    human immunodeficiency virus; HIV; hepatitis B; gene; ds.
XX CLOSTRIDIUM BOTULINUM.
XX
OS US2002107199-A1.
XX
PN 08-AUG-2002.
XX
PD 17-JAN-2002; 2002US-00051952.
XX
PF 05-DEC-2000; 2000US-00730237.
XX
PR (ALLR ) ALLERGAN SALES INC.
XX
PA Walker PS;
XX
PI WPI; 2002-690769/74.
XX
DR
XX
PT Treating a condition e.g., involuntary muscle contraction, spasmodic
    dysphonia, laryngeal dystonia, or wrinkles in an animal or human,
    PT comprises administering a Clostridium neurotoxin component using a
    needleless syringe.
XX
PS Disclosure; Page 15-16; 33pp; English.
XX
XX
CC The invention describes a method of treating a condition in an animal or
    human comprising administering a Clostridium neurotoxin component (1) to
    the subject using a needleless syringe, or administering DNA encoding (1)
    to a cell of the subject in situ. The method is used for treating a
    condition e.g., involuntary muscle contraction or conditions such as
    CC spasmodic dysphonia, laryngeal dystonia, oromandibular dysphonia, lingual
    CC dysphonia, cervical dystonia, focal hand dystonia, blepharospasm,
    CC strabismus, hemifacial spasm, eyelid disorder, cerebral palsy, focal
    CC spasticity, spasmodic colitis, neurogenic bladder, anismus, limb
    CC spasticity, tic, tremor, bruxism, anal fissure, achalasia,
    CC fibromyalgia, dysphagia, wrinkles or brow furrows. The method is also
    CC useful for immunisation against a neurotoxin. By using a needleless
    CC injector, an optimal delivery to a specific tissue layer, for example the
    CC dermis layer is achieved. In a case where the delivery is to the dermis
    CC and not to muscle tissues, the treatment may not cause a loss of motor
    CC function in the area being treated. Also, the use of a needleless
    CC injector, improves clinical safety by eliminating the risk of infection
    CC from accidental injury with needles or from potential splash back of
    CC bodily fluids from liquid jet injectors, thus avoiding the possibilities
    CC of cross-contamination of blood-borne pathogens such as human
    CC immunodeficiency virus (HIV) and hepatitis B. The needleless injector
    CC also offers an optimal and specific delivery of drug particles to treat
    CC conditions with little pain or skin damage such as bruising or bleeding.
    CC This sequence represents the Clostridium botulinum neurotoxin serotype A
    XX
SQ Sequence 3891 BP; 1580 A; 386 C; 635 G; 1290 T; 0 U; 0 Other;

Query Match
Best Local Similarity 50.0%; Score 44; DB 6; Length 3891;
Matches 110; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 1537 ATGTAAGAATTCATAGATGCGGAATAGACGTGATCGATCTTCTGTCACCT 1596
    |||||
DB 2919 ATGGAAGATACCTTAATTAATGTAATTAATCTGACCTTTACAGAGATACAGAAAT 2978
    |||||

QY 1597 ATCATTAAGAAATCATTAAGTAAAGAAACAAACATCAATCTATGTAATTAAGAAACA 1656
    |||||
DB 2979 AAAACAAAGATGATTTTAAATAGAGTCAATGATTAATATATATATATATTAACAG 3038
    |||||

QY 1657 GCTGTTTTCCTCAATTAATTAAGAAAGAAACCATGTGAAGTCAAAATATTTGT 1716
    |||||
DB 3039 ATGGAATTTTGTACTACTCAATTAATTAATGATTAATTAATCTTAATTAATTAATG 3098
    |||||

QY 1717 TTAATCAGGTCATTAAGATCTATTAATTAAGATTTGAAT 1756
    |||||
DB 3099 AGATTAATTAAGATGAACAAATTCATTAATTAAGTAT 3138
    |||||
```

RESULT 20  
 AAD42516  
 ID AAD42516 standard; DNA: 3891 BP.  
 XX  
 AC AAD42516;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Clostridium botulinum toxin A.  
 XX  
 KW Therapy; hyperhidrosis; drug; botulinum toxin; ds.  
 XX  
 OS Clostridium botulinum.  
 XX  
 PN US2002086036-A1.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 05-DEC-2000; 2000US-00730237.  
 XX  
 PR 05-DEC-2000; 2000US-00730237.  
 XX  
 PA (ALLR ) ALLERGAN SALES INC.  
 XX  
 PI Walker PS;  
 XX  
 PS WPI; 2002-618013/66.  
 XX  
 PT Treating hyperhidrosis in a mammal, comprises locally administering a  
 drug particle to an affected skin area without using a needle.  
 XX  
 PS Disclosure; Page 11-13; 29pp; English.  
 XX  
 CC The invention relates to a method for treating hyperhidrosis in a mammal,  
 CC which comprises locally administering a drug particle to an affected skin  
 CC area without using a needle. The method is useful for treating  
 CC hyperhidrosis. The present sequence is Clostridium botulinum toxin used  
 CC in the method of the invention  
 XX  
 SQ Sequence 3891 BP; 1580 A; 386 C; 635 G; 1290 T; 0 U; 0 Other;  
 XX  
 Query Match 2.2%; Score 44; DB 6; Length 3891;  
 Best local Similarity 50.0%; Pred. No. 1.3;  
 Matches 110; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
 QY 1537 ATGTAAGAAATTCATTAGATTGCTGAATAGACTGATCTGTCCTACTTCTGCTCACTT 1596  
 DB 2919 ATGGAAGATTCATTATATATGATGTAATATCTGACCTTACAGGATCTCAGGAAAT 2978  
 QY 1597 ATGATTAAGGAAGTCATTAGCTAAGGAACAAACTCAATCTATNGTATTAGGAACAA 1656  
 DB 2979 AAAACAAAGGATGTTTAAATAGAGTCAAAATGATTATATATACGATTATATAACAG 3038  
 QY 1657 GCTGTTTGGCTCAATATATAAATAAGAAAAAGAAACCATGTAAGAGTCAAAATATTGT 1716  
 DB 3039 ATGATTTTGTACATCTCTATATATATAGATTAAATCACTCAAAATTTATATTAATGS 3098  
 QY 1717 TTAATCAGTCATTGAGATCTATTAAAAAGTAATTGAAAT 1756  
 DB 3099 AAGATTAAATGATGAAAAAACCAATTTCAAATTTAGATTAAT 3138  
 XX  
 RESULT 21  
 ABZ35416/c  
 ID ABZ35416 standard; cDNA: 7093 BP.  
 XX  
 AC ABZ35416;  
 XX  
 DT 05-FEB-2003 (first entry)  
 XX  
 DE Human gene expression profile polynucleotide SEQ ID NO 527.

XX  
 KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
 KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
 KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
 KW gene expression; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200274979-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 20-MAR-2002; 2002WO-US008456.  
 XX  
 PR 20-MAR-2001; 2001US-0276947P.  
 XX  
 PA (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.  
 XX  
 PI Wan J, Wang Y;  
 XX  
 PS WPI; 2002-740862/80.  
 XX  
 PT New gene expression profile generated from primary, endothelial,  
 PT epithelial, and muscle cell types, useful for identifying disease  
 PT pathologies involving alterations of gene expression, e.g. cancer.  
 XX  
 PS Disclosure; Page 667-669; 850pp; English.  
 XX  
 CC The invention relates to a gene expression profile comprising one or more  
 CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type  
 CC is a coronary artery endothelium, umbilical artery or vein endothelium,  
 CC aortic endothelium, dermal microvascular endothelium, pulmonary artery  
 CC endothelium, myometrium microvascular endothelium, keratinocyte  
 CC epithelium, bronchial epithelium, mammary epithelium, prostate  
 CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,  
 CC small airway epithelium, renal epithelium, umbilical artery smooth  
 CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,  
 CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
 CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
 CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,  
 CC osteoblasts or prostate stromal cell. The gene expression profile is used  
 CC for determining the level of RNA expression for a sample, determining the  
 CC phenotype of a cell and distinguishing cell types. The gene or a protein  
 CC expression profile is useful in identifying disease pathologies involving  
 CC alterations of gene expression. The assessment of expression profiles may  
 CC provide meaningful information with respect to tumour type and stage,  
 CC treatment methods, and prognosis. The gene or protein expression profile  
 CC may also be used for creating microarrays. The microarray is useful for  
 CC genetic and physical mapping of genomes, DNA sequencing, genetic or  
 CC medical diagnosis, genotyping of organisms, confirming cell or tissue  
 CC identifications and in identifying promising antibiotics, antiviral or  
 CC antifungal agents  
 XX  
 SQ Sequence 7093 BP; 1999 A; 1487 C; 1453 G; 2154 T; 0 U; 0 Other;  
 XX  
 Query Match 2.2%; Score 44; DB 6; Length 7093;  
 Best local Similarity 63.0%; Pred. No. 1.7;  
 Matches 68; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
 QY 33 GTGAAAAATATATTAACTGTCATATATGTTACAGACCATGATTCGACGACGAAACTC 92  
 DB 6619 GTCAAGTCATCATGTTGTGATATATATGTTGATTCATTTCAAGTAAAGAAACCC 6560  
 QY 93 AATTCAACCAACGTAAGTCAAAAGAAATATATATGCTCATGTAC 140  
 DB 6559 AACTCAAGTAAGTAAAGCAAAAAGGATATTATTGTTCAATTAAC 6512  
 XX  
 RESULT 22  
 AA199260/c  
 ID AA199260 standard; DNA: 10126 BP.  
 XX  
 AC AA199260;

XX 07-JAN-2002 (first entry)  
DT  
XX Human excretory related polynucleotide SEQ ID NO 1024.  
DE  
XX  
XX Human; nootropic; neuroprotective; cytosstatic; dermatological; viroicide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antispasmodic; antianaemic; antiarthritic; cancer;  
KW antihumetic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antileukemic; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;  
KW excretory system; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200155313-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001MO-US001323.  
PF  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218280P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-02331968P.  
PR 14-SEP-2000; 2000US-02332397P.  
PR 14-SEP-2000; 2000US-02332398P.  
PR 14-SEP-2000; 2000US-02332399P.  
PR 14-SEP-2000; 2000US-02332400P.  
PR 14-SEP-2000; 2000US-02332401P.  
PR 14-SEP-2000; 2000US-02333063P.  
PR 14-SEP-2000; 2000US-02333064P.  
PR 14-SEP-2000; 2000US-02333065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.



PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235464P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236372P.  
PR 29-SEP-2000; 2000US-0236377P.  
PR 29-SEP-2000; 2000US-0236378P.  
PR 29-SEP-2000; 2000US-0236386P.  
PR 29-SEP-2000; 2000US-0236389P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249287P.  
PR 17-NOV-2000; 2000US-0249289P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-488784/53.  
XX  
XX New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating and/or preventing human diseases and disorders.  
XX  
XX  
PS Disclosure; SEQ ID NO 925; 564bp + Sequence Listing; English.  
XX  
XX The invention relates to novel kidney related polynucleotides (AA162971-  
CC AA163793) and the encoded polypeptides (AA442417-AA442691) collectively  
CC known as kidney antigens and the use of such kidney antigens for  
CC detecting disorders of the kidney, especially kidney cancer and kidney  
CC cancer metastases. The polynucleotides and proteins are also useful for  
CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC or gene therapy. The genes are isolated from a range of human tissues  
CC disclosed in the specification. The nucleic acids, proteins, antibodies  
CC and (ant)agonists are useful in the diagnosis, treatment and prevention  
CC of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.  
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 10126 BP; 2804 A; 2139 C; 2108 G; 3075 T; 0 U; 0 Other;  
Query Match 2.2%; Score 44; DB 5; Length 10126;  
Best Local Similarity 63.0%; Pred. No. 2;  
Matches 68; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
QY 33 GTGAAATATATTAACGTCATATGTCAGACCATGATTCGACGACGAAC 92  
Db 9649 GTCAAGTCCTCAAGTGTGAGATATATATGTCGATTCATATTCAGTAACGAAACCC 9590  
QY 93 AATTCAACCAACGTAAGTCAAAAGCAAAATATATTTGGCTCATGTAC 140  
Db 9589 AACTCAAGTAAGTAAGTCAAAAGCAAAAGGTAATTTATTTGGTTCACATAC 9542  
RESULT 24  
ADR24206/c  
ID ADR24206 standard; DNA; 5072 BP.  
XX  
XX ADR24206;  
AC  
XX 21-OCT-2004 (first entry)  
DT  
XX  
DE Breast cancer prognosis marker #67.  
XX  
XX db; breast cancer; prognosis; gene expression; diagnosis.  
XX Homo sapiens.  
OS  
XX PN WO2004065545-A2.  
XX  
PD 05-AUG-2004.  
XX  
XX 15-JAN-2004; 2004WO-US001100.  
PF  
XX 15-JAN-2003; 2003US-00342887.  
PR  
XX



PA (ROSE-) ROSETTA INPHARMATICS LLC.  
 PA (NECA-) NETHERLANDS CANCER INST.  
 XX  
 PI Van't Veer LJ, He Y;  
 XX  
 DR WPI; 2004-593473/57.  
 XX  
 PT Classifying a breast cancer patient according to prognosis comprises  
 PT determining the similarity between the level of expression of each of  
 PT five genes in a cell sample taken from patient, to control levels.  
 XX  
 PS Disclosure; SEQ ID NO 67; 226bp; English.  
 XX  
 CC The invention relates to a method of classifying a breast cancer patient  
 CC according to prognosis by determining the similarity between the level of  
 CC expression of each of five genes for which markers are listed in the  
 CC specification, in a cell sample taken from the breast cancer patient, to  
 CC control levels of expression for each respective five genes to obtain a  
 CC patient similarity value. The methods are useful for classifying a breast  
 CC cancer patient according to prognosis. Kits and computer program products  
 CC are useful for data analysis using the diagnostic, prognostic and  
 CC statistical methods of the invention. This sequence corresponds to a  
 CC marker used in the method of the invention.  
 CC  
 SQ Sequence 5072 BP; 1309 A; 1136 C; 1406 G; 1221 T; 0 U; 0 Other;  
 XX  
 XX  
 Query Match 2.2%; Score 43.6; DB 13; Length 5072;  
 Best Local Similarity 50.8%; Pred. No. 1.9;  
 Matches 130; Conservative 0; Mismatches 124; Indels 2; Gaps 1;  
 XX  
 QY 1670 AATATATAAATAAGAAAAAGAAACATGTGAAAGTCAAAATTTGTTAATCAGGTGAT 1729  
 DB 2028 AAACCTTAAAAAAGAAAAAGACAGACTGAAAACTTAATTTTGTGACATATCTAAA 1969  
 QY 1730 TGAATCTATTAAAA--AGTATTGAATCTTTATGATAGAACATCTTGACTCAAGT 1787  
 DB 1968 AAGAGGAGCAAGACAGTACATTTTCTCCCATGGGACAGAAATGAAGGGAGGA 1909  
 QY 1788 GGACAGTGTGAGCTTTTGGCTGTGTCCTACGTAGAAGAGAGCTTTGTCAATAAG 1847  
 DB 1908 GGGAAAGGAGACTTGATTTTCTGTTCACAGAGATGTTGGAGGGGATTTCTTGAG 1849  
 QY 1848 TCTTATATGCTACAGGTGCAAGTTAAGTCCCAAGCTTGCTTAAAGCATACTGAT 1907  
 DB 1848 TTTTTCAGGGAGTGGGACCAAAATGTAGTCCCAAGACTGTCTCAATGATTTGTTGAT 1789  
 QY 1908 TTTGTTTAGACTTTT 1923  
 DB 1788 TTTGTTATAAATCTCT 1773  
 XX  
 RESULT 25  
 ACC85075/C  
 ID ACC85075 standard; DNA; 5073 BP.  
 XX  
 AC ACC85075;  
 XX  
 DT 13-OCT-2003 (first entry)  
 XX  
 DE Human MBCT polypeptide encoding DNA.  
 XX  
 KM MBCT; beta-catenin; cytoskeletal; gene therapy; cancer; human; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 317..1855  
 FT /\*tag= a  
 FT /product= "MBCT"  
 XX  
 XX WO2003052068-A2.  
 XX  
 XX 26-JUN-2003.  
 XX  
 PD

XX  
 PF 12-DEC-2002; 2002WO-US039796.  
 XX  
 PR 13-DEC-2001; 2001US-0340213P.  
 PR 13-DEC-2001; 2001US-0340314P.  
 PR 13-DEC-2001; 2001US-0340322P.  
 PR 15-FEB-2002; 2002US-0357502P.  
 XX  
 PA (EXEL-) EXELIXIS INC.  
 XX  
 PI Costa MA, Gendreau SB, Dora EG, Nicoll M;  
 XX  
 DR WPI; 2003-533010/50.  
 DR P-PSDB; ABR82224.  
 XX  
 PT Identifying a candidate beta-catenin pathway modulating agent for  
 PT diagnosing or treating cancer by detecting a test agent-biased activity  
 PT of the assay system comprising a purified MBCT polypeptide or nucleic  
 PT acid.  
 XX  
 PS Example; Page 57-59; 81pp; English.  
 XX  
 CC The invention relates to genes that modify beta-catenin pathway and to  
 CC the identification of human MBCT (modifiers of beta-catenin) genes. The  
 CC MBCT polypeptides are therapeutic targets for disorders associated with  
 CC beta-catenin function and are useful for manufacturing a medicament for  
 CC diagnosing or treating breast, colon, lung or ovary cancer. The present  
 CC sequence represents a human MBCT polypeptide encoding DNA  
 CC  
 SQ Sequence 5073 BP; 1309 A; 1136 C; 1408 G; 1220 T; 0 U; 0 Other;  
 XX  
 XX  
 Query Match 2.2%; Score 43.6; DB 9; Length 5073;  
 Best Local Similarity 50.8%; Pred. No. 1.9;  
 Matches 130; Conservative 0; Mismatches 124; Indels 2; Gaps 1;  
 XX  
 QY 1670 AATATATAAATAAGAAAAAGAAACATGTGAAAGTCAAAATTTGTTAATCAGGTGAT 1729  
 DB 2029 AAACCTTAAAAAAGAAAAAGACAGACTGAAAACTTAATTTTGTGACATATCTAAA 1970  
 QY 1730 TGAATCTATTAAAA--AGTATTGAATCTTTATGATAGAACATCTTGACTCAAGT 1787  
 DB 1969 AAGAGGAGCAAGACAGTACATTTTCTCCCATGGGACAGAAATGAAGGGAGGA 1910  
 QY 1788 GGACAGTGTGAGCTTTTGGCTGTGTCCTACGTAGAAGAGAGCTTTGTCAATAAG 1847  
 DB 1909 GGGAAAGGAGACTTGATTTTCTGTTCACAGAGATGTTGGAGGGATTTCTTGAG 1850  
 QY 1848 TCTTATATGCTACAGGTGCAAGTTAAGTCCCAAGCTTGCTTAAAGCATACTGAT 1907  
 DB 1849 TTTTTCAGGGAGTGGGACCAAAATGTAGTCCCAAGACTGTCTCAATGATTTGTTGAT 1790  
 QY 1908 TTTGTTTAGACTTTT 1923  
 DB 1789 TTTGTTATAAATCTCT 1774  
 XX  
 RESULT 26  
 AAS46304/C  
 ID AAS46304 standard; DNA; 10369 BP.  
 XX  
 AC AAS46304;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Tumour suppressor gene derived chemically modified sequence #26.  
 XX  
 KM Human; tumour suppressor gene; oncogene; antitumour; cytoskeletal; cancer;  
 KM tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KM cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200168912-A2.  
 XX  
 XX 26-JUN-2003.  
 XX  
 PD

XX 20-SEP-2001.  
PD  
XX  
PF 15-MAR-2001; 2001WO-EP002955.  
XX  
XX 15-MAR-2000; 2000DE-01013847.  
PR 06-APR-2000; 2000DE-01019058.  
PR 07-APR-2000; 2000DE-01019173.  
PR 30-JUN-2000; 2000DE-01032529.  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2001-602752/68.  
XX  
XX Fragments of chemically modified genes associated with tumor suppressor  
PT genes and oncogenes, useful in designing primers and probes for analyzing  
PT diseases associated with cytosine methylation state e.g. cancer.  
XX  
XX Claim 1; SEQ ID NO 26; 27pp; English.  
XX  
XX The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and oncogenes  
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and  
CC 500 are missing from the sequence listing) sequences (Ss) and sequences  
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-  
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of  
CC probes for detecting the cytosine methylation state and/or single  
CC nucleotide polymorphisms and also to be used in an array for analysing  
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The  
CC probes can also be used in a method for ascertaining genetic and/or  
CC epigenetic parameters for the diagnosis and/or therapy of existing  
CC diseases or the predisposition to specific diseases, by analysing  
CC cytosine methylations. The parameters may be compared to another set of  
CC genetic and/or epigenetic parameters, the differences serving as basis  
CC for diagnosis and/or prognosis events which are disadvantageous to  
CC patients. The present sequence is one of the 533 genomic sequences  
CC derived from tumour suppressor genes and oncogenes. Sequences with even  
CC numbered Seq ID numbers are the complementary sequence of the  
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID  
CC 535, except for those whose partner sequence is missing). Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WPIO  
CC at ftp.wpiio.int/pub/published\_pct\_sequences  
XX  
XX Sequence 10369 BP; 2628 A; 319 C; 2617 G; 4805 T; 0 U; 0 Other;  
SQ  
Query Match 2.2%; Score 43.6; DB 4; Length 10369;  
Best Local Similarity 51.5%; Pred. No. 2.6;  
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
QY 1581 ATTTCTGCTGCTCACTATGATAGAGAGTCAAGTGAAGCAAACTACATCTAT 1640  
DB 8333 ATTATCTCTTAATTTATTAATCTAAACATTAACCTAAATTAATCTAACCACTTA 8274  
QY 1641 GTAATTAGAGAAACAAGCTGTTTGGCTCAATTAATAAATAAGAAAGCAATGTGA 1700  
DB 8273 TCAGAAACATTAATCAACTTTTATACAAAAAATTAACAAATTTTAACATAACAT 8214  
QY 1701 AAGTCAAAATATTTGTTTAATCAAGTCATGAGATCTATTAAGATTTGAATTCCT 1760  
DB 8213 AATTAATACATACCTATATATCCCACTACTCGAAAAAATAAACAATTAATCTTAACCTTA 8154  
QY 1761 TATGATGAGAACTA 1774  
DB 8153 AAAAAGCAAAACTA 8140

RESULT 27  
AB132393/c

ID ABL32393 standard; DNA; 10369 BP.  
XX  
XX ABL32393;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX  
DE Human immune system associated gene SEQ ID NO: 366.  
XX  
XX Human immune system associated gene  
XX Human; immune system disease; cytosine methylation; antiaesthetic;  
KW antiarteriosclerotic; antihaemic; cytosolic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antineumatic; antiaesthetic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
KW ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200200928-A2.  
PN  
XX  
XX 03-JAN-2002.  
PD  
XX  
XX 02-JUL-2001; 2001WO-BP007537.  
PF  
XX  
PR 30-JUN-2000; 2000DE-01032529.  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.  
XX  
XX Claim 1; SEQ ID NO 366; 32pp + Sequence listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention  
XX  
XX Sequence 10369 BP; 2628 A; 319 C; 2617 G; 4805 T; 0 U; 0 Other;  
SQ  
Query Match 2.2%; Score 43.6; DB 6; Length 10369;  
Best Local Similarity 51.5%; Pred. No. 2.6;  
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
QY 1581 ATTTCTGCTGCTCACTATGATAGAGTCAAGTGAAGCAAACTACATCTAT 1640  
DB 8333 ATTATCTCTTAATTTATTAATCTAAACATTAACCTAAATTAATCTAACCACTTA 8274  
QY 1641 GTAATTAGAGAAACAAGCTGTTTGGCTCAATTAATAAATAAGAAAGCAATGTGA 1700  
DB 8273 TCAGAAACATTAATCAACTTTTATACAAAAAATTAACAAATTTTAACATAACAT 8214  
QY 1701 AAGTCAAAATATTTGTTTAATCAAGTCATGAGATCTATTAAGATTTGAATTCCT 1760  
DB 8213 AATTAATACATACCTATATATCCCACTACTCGAAAAAATAAACAATTAATCTTAACCTTA 8154  
QY 1761 TATGATGAGAACTA 1774  
DB 8153 AAAAAGCAAAACTA 8140

RESULT 28

AD85280/c  
ID AD85280 standard; DNA; 1212 BP.  
XX  
AC AD85280;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Aspergillus fumigatus essential gene with introns #504.  
XX  
KM Fungicide: Aspergillus fumigatus infection; Farmer's lung disease;  
KW drug screening; ds.  
OS Aspergillus fumigatus.  
XX  
PN WO2004067709-A2.  
XX  
PD 12-AUG-2004.  
XX  
PF 16-JAN-2004; 2004WO-US001099.  
XX  
PR 17-JAN-2003; 2003US-0441281P.  
PR 13-JUN-2003; 2003US-0478196P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
PA (ELIT-) ELITRA CANADA LTD.  
XX  
PI Jiang B, Hu W, Lemieux S, Roemer T;  
XX  
DR WPI; 2004-594200/57.  
DR P-PSDB; AD86454.  
XX  
PT New purified or isolated Aspergillus fumigatus nucleic acid molecule  
PT encoding a gene product, useful for diagnosing and/or treating invasive  
PT fungal infections, such as Farmer's lung disease.  
XX  
PS Claim 3; SEQ ID NO 1504; 164pp; English.  
XX  
CC The present invention relates to Aspergillus fumigatus genes that are  
CC essential and are potential targets for drug screening. The methods and  
CC compositions of the present invention are useful for diagnosing and/or  
CC treating invasive Aspergillus fumigatus infection, including the allergic  
CC forms of the disease, such as Farmer's lung disease. They can also be  
CC used in various drug discovery purposes, such as expression of the  
CC recombinant protein, hybridization assay and construction of nucleic acid  
CC arrays. The present sequence represents an Aspergillus fumigatus  
CC essential gene sequence containing all introns, used during diagnosis and  
CC drug development in the invention. These genes share a high degree of  
CC sequence conservation with known essential genes of candida albicans. The  
CC sequence data for this patent is not represented in the printed  
CC specification, but was obtained in electronic format from WIPO.  
XX  
XX  
SQ Sequence 1212 BP; 311 A; 328 C; 335 G; 238 T; 0 U; 0 Other;  
Query Match 2.2%; Score 43.4; DB 13; Length 1212;  
Best Local Similarity 50.7%; Pred. No. 1.2;  
Matches 104; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 163 GGAAGGGGCTTTGGGAACAAGAAATGTTCTCAAAATTCAGAAATCTAGATTATGTC 222  
DB 756 GGAAGATATTTGGGCAAGAATCTTGTGGCCAAATTCAGACCACTTGTCTTCGGC 697  
QY 223 AGATGGGTACCTTCTGTCCCTGAGGTGGTGTAGCGATGTAGATCTTATGGAGG 282  
DB 696 TGAATGCTGTTCAACCATATATGTGTGTAGCGGAACGACAGATGTGAATTCGGAGG 637  
QY 283 AAAGAGTCATGTTAGATGAAGTAGGCTTAAGCAAAAGGGCAAGGCCACTATATC 342  
DB 636 AAGAGTAAAGGTTTGGCGAGTTCTAGGGAATGGGACCAAGATTAGATGAACATAGT 577  
QY 343 ATGCTAAATATGTTTTTTTGTATG 367  
DB 576 CGCCTAAAAATGTTGTAACCTTACTG 552

RESULT 29  
ADZ70960/c  
ID ADZ70960 standard; DNA; 1236 BP.  
XX  
AC ADZ70960;  
XX  
DT 14-JUL-2005 (first entry)  
XX  
DE Human chromosome 1 contig DNA SEQ ID NO 84.  
XX  
KM ds; matrix attachment region; MAR; protein production.  
XX  
KW Homo sapiens.  
OS  
XX  
PN WO2005040377-A2.  
XX  
PD 06-MAY-2005.  
XX  
PF 22-OCT-2004; 2004WO-EP011974.  
XX  
PR 24-OCT-2003; 2003US-0513574P.  
PR 06-FEB-2004; 2004EP-00002722.  
XX  
PA (SELE-) SELEXIS SA.  
XX  
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;  
PI Putini S;  
XX  
DR WPI; 2005-333507/34.  
XX  
PT New purified and isolated DNA sequence having protein production  
PT increasing activity comprises a bent DNA element and a binding site for a  
PT DNA binding protein, useful for increasing protein production activity in  
PT eukaryotic host cell.  
XX  
PS Disclosure; SEQ ID NO 84; 282pp; English.  
XX  
CC The invention relates to a purified and isolated DNA sequence having  
CC protein production increasing activity comprising at least one bent DNA  
CC element, and at least one binding site for a DNA binding protein. The  
CC purified and isolated DNA sequence comprising a first and a second  
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a  
CC MAR nucleotide sequence selected from a purified and isolated DNA  
CC sequence above, a purified and isolated MAR DNA above, a purified and  
CC isolated cysMAR element and/or fragment, a synthetic MAR sequence, a  
CC sequence complementary to it, its molecular chimera, or its combinations  
CC and variants, is useful for increasing protein production activity in a  
CC eukaryotic host cell. The present sequence represents a human chromosome  
CC 1 contig DNA.  
XX  
XX  
SQ Sequence 1236 BP; 506 A; 28 C; 126 G; 576 T; 0 U; 0 Other;  
Query Match 2.2%; Score 43.4; DB 14; Length 1236;  
Best Local Similarity 45.1%; Pred. No. 1.2;  
Matches 161; Conservative 0; Mismatches 196; Indels 0; Gaps 0;  
QY 1408 TATATCATATCATATATGAGAGCTGCTTTTAAAGCGTAATGTTTATATGACTAAG 1467  
DB 397 TATATCATATCAT 338  
QY 1468 GCTTTGGCTTACATATATTAAGGGGATTTAGCAATGTGATACAGAGTCTTTTCTCAC 1527  
DB 337 TATATCAT 278  
QY 1528 AGCTTCATATGTAAGAAATTCATTTAGATTGGCTGAATAGACTGATCTGTCATTCTC 1587  
DB 277 AT 218  
QY 1588 TGCTCACTATCATATAGAGATGATTAAGTAAGAAACAAATCTCAATCTATGTATATA 1647  
DB 217 AT 158

Oy 1648 GAAGAACACGCGGTTTGTCATATATAAATAGAAAAAACAATGGAAGCTCA 1707  
 Db 157 ATATATATCATTTATATATCTATATATATATATATATATATATATATAT 98  
 Oy 1708 AATATTTGTTAATACGCTCATTTGAAATCTATTTAAAAAGTATTTGAATCTTATATG 1764  
 Db 97 AATG 41

RESULT 30  
ADR84693/

ID ADR84693 standard; DNA; 3212 BP.

AC ADR84693;

DT 04-NOV-2004 (first entry)

DE *Aspergillus fumigatus* essential gene genomic sequence #504.

KW Fungicide; *Aspergillus fumigatus* infection; Farmer's lung disease;

drug screening; ds.

05 *Aspergillus fumigatus*.

PN WO2004067709-A2.

PD 12-AUG-2004.

PF 16-JAN-2004; 2004WO-US001099.

PR 17-JAN-2003; 2003US-0441281P.

PR 13-JUN-2003; 2003US-0478196P.

PA (ELIT-) ELITRA PHARM INC.

PA (ELIT-) ELITRA CANADA LTD.

PI Jiang B, Hu W, Lemieux S, Roemer T;

DR WPI; 2004-594200/57.

XX

PT New purified or isolated *Aspergillus fumigatus* nucleic acid molecule

PS Claim 3; SEQ ID NO 504; 164pp; English.

CC The present invention relates to *Aspergillus fumigatus* genes that are  
CC essential and are potential targets for drug screening. The methods and  
CC compositions of the present invention are useful for diagnosing and/or  
CC treating invasive *Aspergillus fumigatus* infection, including the allergic  
CC forms of the disease, such as Farmer's lung disease. They can also be  
CC used in various drug discovery purposes, such as expression of the  
CC recombinant protein. hybridization assay and construction of nucleic acid  
CC arrays. The present sequence represents an *Aspergillus fumigatus*  
CC essential gene full length genomic sequence, used during diagnosis and  
CC drug development in the invention. These genes share a high degree of  
CC sequence conservation with known essential genes of candida albicans. The  
CC sequence data for this patent is not represented in the printed  
CC specification, but was obtained in electronic format from WIPO.  
XX  
XX Sequence 3212 BP; 870 A; 810 C; 762 G; 770 T; 0 U; 0 Other;

Query Match	2.2%	Score 43.4	DB 13	Length 3212
Best Local Similarity	50.7%	Pred. No. 1.8		
Matches 104, Conservative	0	Mismatches 101	Indels 0	Gaps 0

OY 163 GGAAGGGCTTTGGGAACAAGATTTGTTCTCAATTCTAGGATCTACGATTAGTCC 222  
 Db 1756 GGAAGATATTGCGGACACAGATCTTGTGTGCACATTTCGAGACACACCTTCTCTCGGC 1697  
 OY 223 AGGATGGGTCACCTTCTGTCTCCGAGGTGCTGTAGGATGCTTATGGGAGG 282

Db 1596 TGAATGCTGTTCACCCCATATATATGCTGTACGGGAACGACGAGATGTGAATTTCCGGAGG 1637

Qy 283 AAAGATGTCAGTCTTTAGGATAGAGTGGGCTTAAGCAACAAGGCGAAGGGCCCTATATC 342

Db 1536 AAGAGTAAAGGTTTGGCGAGTTCTTAAAGCAATGGGACCCAGATTAGCATTAACCACTAGT 1577

Qy 343 ATGCTAAAAAATGCTTTTGTGATG 367

Db 1576 CGCCTAAAAATGTTGTAACTTACG 1552

### RESULT 31

ID	standard; DNA; 32216 BP.
AAL06925	

AC AAL06925 ;

DT 21-NOV-2001 (first entry)

DE Human reproductive system related antigen DNA SEQ ID NO: 9613.

XX  
XX

KW cancer; gene therapy; ds.

OS Homo sapiens.

PN WO200155320-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US001339.

XX

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 16-MAR-2000; 2000US-0189874P.

PR 18-APR-2000; 2000US-0198123P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.

PR 11-JUL-2000; 2000US-0217487P.

PR	14-JUL-2000;	2000US-0218290P.
PR	26-JUL-2000;	2000US-0220963P.
PR	26-JUL-2000;	2000US-0220964P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225447P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.

PR 30-AUG-2000; 2000US-0228974P.  
PR 01-SEP-2000; 2000US-0229267P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.



KM neurological disease; infection; human; secreted protein; ds.  
XX Homo sapiens.  
XX OS  
XX PN WO200155325-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001345.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0180664P.  
PR 02-MAR-2000; 2000US-0186330P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241212P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-488786/53.  
 DR  
 XX New isolated ovarian and/or breast cancer related nucleic acids and  
 PT polypeptides, useful for diagnosing, treating and/or preventing human  
 PT diseases and disorders, particularly ovarian and/or breast cancer.  
 XX  
 PS Disclosure; SEQ ID NO 536; 577bp + Sequence listing; English.  
 XX  
 CC The invention relates to novel genes (ABA07454-ABA08224) and proteins  
 CC (AB10743-AB10980) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX  
 SQ Sequence 32216 BP; 8496 A; 6430 C; 6479 G; 10811 T; 0 U; 0 Other;  
 Query Match 2.1%; Score 43; DB 4; Length 32216;  
 Best Local Similarity 55.8%; Pred. No. 6;  
 Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
 QY 1653 ACAAGCTGTTTCTCAATATATAAAGAAAGAAACATGGAAGTCAAAATAT 1712  
 DB 23479 ACAGGTGAGATTCATCTCAAAAATAAATAAATAAATAAATTAATCTGAAAATTC 23420  
 QY 1713 TTGTTTAATCAGCTCATTTGAAATCTATTAATAAAGTATTTGAATCTTTATGATGAGAC 1772  
 DB 23419 TGATTAAGAGAGACATCTTAATCTATTTAGAAAATTTTAATTTGTGATTTTAAAT 23360  
 QY 1773 TATCTGACCTCAAGTGAGACAGTGATG 1799  
 DB 23359 TTTACTACCTGAATCAAAAATTTGGTGA 23333  
 RESULT 33  
 ADF51132  
 ID ADF51132 standard; DNA; 243428 BP.  
 XX  
 AC ADF51132;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX  
 DE Human P-Rex1 genomic DNA sequence.  
 XX  
 KW human; P-Rex1; Rac; guanine-nucleotide exchange factor; GEF; GTPase;  
 KW inflammation; metastasis; septic shock; neurodegeneration;  
 KW atherosclerosis; antiinflammatory; cytosolic; antibacterial;  
 KW immunosuppressive; neuroprotective; antiarteriosclerotic; gene; ds.  
 OS Homo sapiens.  
 XX  
 PN MO2003080664-A1.  
 XX  
 PD 02-OCT-2003.  
 XX  
 PF 21-MAR-2003; 2003WO-GB001238.  
 XX

PR 21-MAR-2002; 2002GB-00006684.  
 XX  
 PA (BABR-) BABRAHAM INST.  
 XX  
 PI Stephens L, Hawkins PT;  
 DR WPI; 2004-011515/01.  
 DR P-PSDB; ADF51119.  
 XX  
 PT New isolated P-Rex1 protein or its derivative useful for discovering  
 PT drugs capable of reducing or inhibiting inflammation, metastasis, septic  
 PT shock, neurodegeneration or atherosclerosis, or for identifying P-Rex1  
 PT modulators.  
 XX  
 PS Disclosure; SEQ ID NO 14; 198bp; English.  
 XX  
 CC This invention relates to a novel protein useful as an anti-inflammatory  
 CC target. Specifically, it refers to the guanine-nucleotide exchange factor  
 CC (GEF) named P-Rex1, which has also been identified as a  
 CC phosphatidylinositol (3,4,5)P3-sensitive activator of Rac (a monomeric  
 CC GTPase). Accordingly, P-Rex1 can be described as having Rac-GEF activity  
 CC and is adapted to function downstream of activation of heterotrimeric G  
 CC proteins in neutrophils. The present invention describes this protein as  
 CC a useful target for drug discovery or for discovery of a drug capable of  
 CC reducing or inhibiting inflammation, metastasis, septic shock,  
 CC neurodegeneration or atherosclerosis. As such, P-Rex1 can have various  
 CC activities including antiinflammatory, cytosolic, antibacterial,  
 CC immunosuppressive, neuroprotective and antiarteriosclerotic. Furthermore,  
 CC the protein or its mutant, the nucleic acid or appropriate antibody may  
 CC be used in a screening assay to identify a modulator of P-Rex1 binding,  
 CC activity or expression. This polynucleotide is the human P-Rex1 genomic  
 CC DNA sequence of the invention.  
 CC  
 XX  
 SQ Sequence 243428 BP; 65880 A; 63219 C; 59010 G; 55319 T; 0 U; 0 Other;  
 Query Match 2.1%; Score 43; DB 12; Length 243428;  
 Best Local Similarity 59.3%; Pred. No. 14;  
 Matches 73; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 75 TCGAAGTGACAGAACTCAATCAACCAAGCTAAAGTCAAAAGAAATATATTTGGCTCA 134  
 DB 238428 TCGAAGTGACAGAAATCAACTCAACCAAGCTCAAGGAAATATATTTGGCTCA 238487  
 QY 135 TGTAACTTCTCAAGAGAGGCGAGAGTGAAGGGGCTTTGGAAACAAGAAATTTGCT 194  
 DB 238488 TCTAAGCTGTCTAATCTTGGGAGATGCGTGAATCCAGAGTCTCAAGATTTGCTTCTCC 238547  
 QY 195 CAA 197  
 DB 238548 CAA 238550  
 RESULT 34  
 ABL32695/C  
 ID ABL32695 standard; DNA; 9963 BP.  
 XX  
 AC ABL32695;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 668.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytosolic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriasis; anaemia;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; acute  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
 ds.  
 OS Homo sapiens.  
 XX  
 XX



PN WO200200928-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP007537.  
XX  
XX 30-JUN-2000; 2000DE-01032529.  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPIC-) EPIDEMIOLOGICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX MPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.  
XX  
XX  
PS Claim 1; SEQ ID NO 668; 32pp + Sequence listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention  
XX  
XX  
SQ Sequence 9963 BP; 2584 A; 180 C; 2281 G; 4918 T; 0 U; 0 Other;  
  
Query Match 2.1%; Score 42.8; DB 6; Length 9963;  
Best Local Similarity 52.2%; Pred. No. 4.1;  
Matches 95; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
  
QY 1596 TATCATAGGAAGTCTAGCTAAGACAAACATCATCTATGTAATTAGAGACG 1655  
DB 7988 TATTCACAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7929  
QY 1656 AGCTGTTTGTCTCATATATAAATAAAGAAACATGGAAGTCAAAATATTTG 1715  
DB 7928 ATAAAAAACCATATCATCTCAATATAAATAAATAAATAAATAAATAAATAA 7869  
QY 1716 TTTAATCAGCTCATGTGAAATCTATTAAAAAGATTGTAATTCATGAGAACTAT 1775  
DB 7868 ATTAATTAATAATTAATAAATCAATCTCTTATTAAAAATTTTAAATAATTAATTAAT 7809  
QY 1776 CT 1777  
DB 7808 AT 7807  
  
RESULT 35  
Continuation (33 of 34) of AEB42401 from base 3200001 (L. pneumophila DNA SEQ ID NO 6733  
WP Sequence Split into 34 fragments LOCUS AEB42401 Accession Aeb42401  
WP Fragment Name Begin End  
WP AEB42401\_00 1 110000  
WP AEB42401\_01 100001 210000  
WP AEB42401\_02 200001 310000  
WP AEB42401\_03 300001 410000  
WP AEB42401\_04 400001 510000  
WP AEB42401\_05 500001 610000  
WP AEB42401\_06 600001 710000  
WP AEB42401\_07 700001 810000  
WP AEB42401\_08 800001 910000  
WP AEB42401\_09 900001 1010000  
WP AEB42401\_10 1000001 1110000  
WP AEB42401\_11 1100001 1210000  
WP AEB42401\_12 1200001 1310000  
WP AEB42401\_13 1300001 1410000  
WP AEB42401\_14 1400001 1510000

WP AEB42401\_15 1500001 1610000  
WP AEB42401\_16 1600001 1710000  
WP AEB42401\_17 1700001 1810000  
WP AEB42401\_18 1800001 1910000  
WP AEB42401\_19 1900001 2010000  
WP AEB42401\_20 2000001 2110000  
WP AEB42401\_21 2100001 2210000  
WP AEB42401\_22 2200001 2310000  
WP AEB42401\_23 2300001 2410000  
WP AEB42401\_24 2400001 2510000  
WP AEB42401\_25 2500001 2610000  
WP AEB42401\_26 2600001 2710000  
WP AEB42401\_27 2700001 2810000  
WP AEB42401\_28 2800001 2910000  
WP AEB42401\_29 2900001 3010000  
WP AEB42401\_30 3000001 3110000  
WP AEB42401\_31 3100001 3210000  
WP AEB42401\_32 3200001 3310000  
WP AEB42401\_33 3300001 3345687  
  
Query Match 2.1%; Score 42.8; DB 14; Length 110000;  
Best Local Similarity 46.9%; Pred. No. 11;  
Matches 134; Conservative 0; Mismatches 152; Indels 0; Gaps 0;  
  
QY 1531 TCTCATATGTAAGAATTCATAGATTGGCTGAATAGACTGATCTGTCATTCTCTGC 1590  
DB 82357 TCTATTATGCTGGAAATTCATTTTGGCAACATGATTTTGTGTCATTAAATCGCC 82298  
QY 1591 TCACCTATCATAGAAGTCAATAGCTAAGGAAACAAACATCAATCTATGTAATTAGAA 1650  
DB 82297 TTTTCATATCCAAACATAGGCTATCTTTTGAAGAAGACGACTAAAAAGATTGGATTAAAA 82238  
QY 1651 GAACAAGCTGTTTGTCTCAATATAAATAAATAAAGAAACCATGTGAAAGTCAAAT 1710  
DB 82237 ATATGATTTGGGGAGTACTCGGTTTCAGAAATAATATCTGTCAAAATGACAAATATAAT 82178  
QY 1711 ATTGTGTTAATCAGCTCATTTGGAATCTATTAATAAGTATTGAATCTTTATGATGAG 1770  
DB 82177 TTTCCGTTAATATGTCATAGTGTGTTTTTAAAGGATCTATCAAAATTTTAAAGGAGTAAA 82118  
QY 1771 ACTATCTTGACTCAAGTGCAGAGTGTGAGCTTTTGGCTGTGT 1816  
DB 82117 AGCAGTATGATTTAATTCACGCTGTAAAGCTTATGCTTTGGGT 82072  
  
RESULT 36  
ABL33091/c  
ID ABL33091 standard; DNA; 16724 BP.  
XX ABL33091;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX  
DE Human immune system associated gene SEQ ID NO: 1064.  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antineumatic; antiarthritic; antidiabetic; antipsoriasis;  
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200200928-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP007537.  
XX  
XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.  
 XX (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful for  
 PT diagnosis and treatment of diseases associated with abnormal cytosine  
 PT methylation.  
 XX  
 PS Claim 1; SEQ ID NO 1064; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention  
 XX  
 SQ Sequence 16724 BP; 5336 A; 237 C; 3558 G; 7593 T; 0 U; 0 Other;  
 XX  
 Query Match 2.1%; Score 42.6; DB 6; Length 16724;  
 Best Local Similarity 50.2%; Pred. No. 5.8;  
 Matches 135; Conservative 0; Mismatches 129; Indels 5; Gaps 1;  
 QY 1517 CTTTCTCCACAGCTCATATGTAAGAATTCATTAGATGGCTGAATAGACTGATCT 1576  
 DB 16680 CCTTACAAACATATAATATATATATATATATATATATATATATATATATATAT 16621  
 QY 1577 GTCCATTTCTGCTCACTATATCATAGAGTCATTAGCTAAGAACAAATCTACAT 1636  
 DB 16620 CCTAAATCTATTAATCTTAAATATCAAAAACAATTAATTAATTCCTAAAAAATC 16561  
 QY 1637 CTATGTAATTAGAAACAAGCTGTTTGTCTCAATTAATAAATAAGAAACAT 1696  
 DB 16560 TTTCTTAAACAACTAACGCTTAATACAAACAATTAATAAATAAATAAACAACAC 16501  
 QY 1697 GTGAAGTCAAAATATTTGTTTATATCAGTCATTGAGATCTATTAAAGTATTTGAAT 1756  
 DB 16500 TTTAAACATTTAT 16446  
 QY 1757 TCTTTATGATGAGACTATCTGACTCAA 1785  
 DB 16445 TAATTACAAACAACTAATATATATATATATATATATATATATATATATATATAT 16417  
 XX  
 RESULT 37  
 ABL34537/c  
 ID ABL34537 standard; DNA; 16724 BP.  
 XX  
 AC ABL34537;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human metastasis associated gene SEQ ID NO: 90.  
 XX  
 KW Metastasis associated gene; cytosine methylation; gene therapy; cancer;  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200177376-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-EP003970.  
 XX  
 PR 06-APR-2000; 2000DE-01019058.  
 XX  
 PR 07-APR-2000; 2000DE-01019173.  
 XX

PR 30-JUN-2000; 2000DE-01032529.  
 XX 01-SEP-2000; 2000DE-01043826.  
 XX  
 PA (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-010922/01.  
 XX  
 PT New nucleic acid derived from chemically treated metastasis genes, useful  
 PT for diagnosis of cancers by analysis of cytosine methylation, also for  
 PT treatment.  
 XX  
 PS Claim 1; SEQ ID NO 90; 23pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of human metastasis associated  
 CC genes which are modified by cytosine methylation. The sequences can be  
 CC used in the diagnosis and treatment of cancer. The present sequence is  
 CC one of the genes of the invention  
 XX  
 SQ Sequence 16724 BP; 5336 A; 237 C; 3558 G; 7593 T; 0 U; 0 Other;  
 XX  
 Query Match 2.1%; Score 42.6; DB 6; Length 16724;  
 Best Local Similarity 50.2%; Pred. No. 5.8;  
 Matches 135; Conservative 0; Mismatches 129; Indels 5; Gaps 1;  
 QY 1517 CTTTCTCCACAGCTCATATGTAAGAATTCATTAGATGGCTGAATAGACTGATCT 1576  
 DB 16680 CCTTACAAACATATAATATATATATATATATATATATATATATATATATATAT 16621  
 QY 1577 GTCCATTTCTGCTCACTATATCATAGAGTCATTAGCTAAGAACAAATCTACAT 1636  
 DB 16620 CCTAAATCTATTAATCTTAAATATCAAAAACAATTAATTAATTCCTAAAAAATC 16561  
 QY 1637 CTATGTAATTAGAAACAAGCTGTTTGTCTCAATTAATAAATAAGAAACAT 1696  
 DB 16560 TTTCTTAAACAACTAACGCTTAATACAAACAATTAATAAATAAATAAACAACAC 16501  
 QY 1697 GTGAAGTCAAAATATTTGTTTATATCAGTCATTGAGATCTATTAAAGTATTTGAAT 1756  
 DB 16500 TTTAAACATTTAT 16446  
 QY 1757 TCTTTATGATGAGAACTATCTTACTCAA 1785  
 DB 16445 TAATTACAAACAACTAATATATATATATATATATATATATATATATATATATAT 16417  
 XX  
 RESULT 38  
 ABL70260/c  
 ID ABL70260 standard; DNA; 16724 BP.  
 XX  
 AC ABL70260;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Chemically treated cell signalling DNA sequence complementary to#75.  
 XX  
 KW Cell signalling; cytosine methylation; cell signalling disease; cancer;  
 XX  
 OS Unidentified.  
 XX  
 PN WO200202807-A2.  
 XX  
 PD 10-JAN-2002.  
 XX  
 PF 29-JUN-2001; 2001WO-EP007471.  
 XX  
 PR 30-JUN-2000; 2000DE-01032529.  
 XX  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX  
 PA (EPiG-) EPIGENOMICS AG.  
 XX



Oy	1577	GTCCATTTCTCCTGCTACCTATCATTAAGAAGATCATTAGCTAAGAAACAAAACATTCACAT	1636
Dd	16620	CCCTAAATCTCTTAACCTCTTAATAATACAAAACATTAATATTAATCTCTTAATAAAATTC	16561
Oy	1637	CTATGTAATTAAGAAACAAGCTGGTTTCTCCTCATATTAATAAATAAGAAAAACAT	1696
Dd	16560	TTTCTTAAAAACAACATACGCCCTAATATACAAACATATTAATAAATAAACAACACAC	16501
Oy	1697	GTGAAGACGAAATATTTGTTTAACTCGGTCATTTGAGAAATCTATTAAAAAGTTTGAAT	1756
Dd	16500	TTTAAAAACATTAATATTAATTAATTAATTAACA-----AAAAATAAAAAATTATAAAAAA	16468
Oy	1757	TCTTATGATGAGAACTATCTGACTCA	1785
Dd	16445	TAAATTACAAACAAACTAATATTATCCCA	16417
RESULT 40			
ABX39419			
XX	AC	ABX39419 standard; cDNA; 416 BP.	
XX	XX	ABX39419;	
XX	DT	20-FEB-2003 (first entry)	
XX	DE	Bovine EST associated with lactation/muscle/fat deposition #4584.	
XX	KW	Bovine; ss; EST; expressed sequence tag; lactation; LMPD;	
XX	KM	muscle deposition; fat deposition; genome mapping; gene identification;	
XX	KW	gene analysis; cattle breeding.	
OS	Bos	Taurus.	
XX	PN	US2002137139-A1.	
XX	PD	26-SEP-2002.	
XX	PF	24-SEP-2001; 2001US-00960352.	
XX	PR	12-JAN-1999; 99US-0115707P.	
XX	PR	11-JAN-2000; 2000US-00480902.	
XX	PA	(BYAT/) BYATT J C.	
XX	PA	(MATH/) MATHIALAGAN N.	
XX	PA	(TAON/) TAO N.	
XX	PA	(WARR/) WARREN W C.	
XX	PI	Byatt JC, Mathialagan N, Tao N, Warren WC;	
XX	PI	WPI; 2003-110599/10.	
XX	PT	New nucleic acid associated with lactation, and muscle and fat	
XX	PT	deposition, useful for genome mapping, gene identification and analysis,	
XX	PT	cattle breeding, or for genetically improving cattle.	
XX	PS	Claim 2; SEQ ID NO 4584; 245bp; English.	
XX	XX	The invention relates to a purified nucleic acid molecule associated with	
XX	XX	lactation or muscle and fat deposition (designated LMPD), derived from	
XX	XX	cattle, and the LMPD nucleic acid can specifically hybridise to a second	
XX	XX	nucleic acid molecule comprising any of 15112 nucleotide sequences,	
XX	XX	appearing as ABX34836-ABX49947, or complements of them. Also included are	
XX	XX	; (1) a transformed cell having a nucleic acid comprising an LMPD nucleic	
XX	XX	acid linked to a promoter and a 3' non-translated sequence that	
XX	XX	functions in the cell to cause termination of transcription and addition	
XX	XX	of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and	
XX	XX	(2) determining a level or pattern of a molecule in a bovine cell or	
XX	XX	tissue comprising: (a) incubating a marker nucleic acid (comprising any	
XX	XX	of the 15112 nucleic acid sequences or its complement or fragment) with a	
XX	XX	complementary nucleic acid molecule obtained from the bovine cell or	
XX	XX	tissue, where hybridisation between the marker nucleic acid and the	
XX	XX	complementary nucleic acid permits the detection of the molecule; and (b)	
XX	XX	detecting the level or pattern of the complementary nucleic acid, where	

CC	the detection of the complementary nucleic acid is predictive of the
CC	level or pattern of the molecule. The LMPD nucleic acid is used for
CC	determining a level or pattern of a molecule in a bovine cell or tissue.
CC	It is useful for genome mapping, gene identification and analysis, cattle
CC	breeding, preparation of constructs for use in cattle gene expression, or
CC	for genetically improving cattle. The present sequence is one of the
CC	15112 bovine LMPD BSR (expressed sequence tag) nucleic acids. Note: The
CC	present sequence was not shown in the specification but was obtained in
CC	electronic format from the USPTO web site:
CC	seqdata.uspto.gov/sequence.html?docid=20020137139
XX	
SQ	Sequence 416 BP; 223 A; 1 C; 14 G; 178 T; 0 U; 0 Other;
Query Match	2.1%; Score 42.4; DB 8; Length 416;
Best Local Similarity	52.9%; Pred. No. 1.4;
Matches	91; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
OY	1609 TCATTAGCTAAGAAACAAAATCATCTATGTGAATTAGAGAACAGCTGTTTGGCT 1668
DB	134 TTATTAATAAAAAAAAAATTAATTTAAAAAAAAAAAAAAAAAAAAAAGTTTTAA 193
OY	1669 CAATATTAATAATAAGAAAAACAACATGTGAAGTCGAAATATTTGTTAATCAGGTCA 1728
DB	194 AAATTAATAATAATAATAATTAATAAATTCATBAATAAAAAAAAAAATTTTAAAAAATTT 253
OY	1729 TTGAGAATCTATTAATAAAGTATTTGAAATCTTTATATGATGGAACACTATCTTGA 1780
DB	254 TTAAATTTTAAAGAAAAAATTTTAAAAAATTTTAAATTTATTAATAATTATAGA 305
XX	
RESULT 41	
AAZ87220	
ID	AAZ87220 standard; DNA; 1317 BP.
XX	
AC	AAZ87220;
XX	
DT	08-MAY-2000 (first entry)
DE	DNA encoding native BoNT serotype A (BONTA) C-terminal fragment (Hc).
XX	
KW	Boculinum neurotoxin; heavy chain; BoNT; serotype A; C-terminal fragment;
KM	Hc; Venezuelan equine encephalitis virus replicon; VEE; botulinm;
XX	vaccine; diagnosis; drug screening; ds.
OS	Clostridium botulinum.
FH	Key
FT	CDs
FT	1 1317
FT	/tag= a
FT	/product= "Native botulinum neurotoxin serotype A (BONTA)
FT	heavy chain C-terminal fragment"
FT	/transl_except= (pos:1294..1314, aa:GLY)
XX	
PN	WO200002524-A2.
PD	20-JAN-2000.
XX	
PF	09-JUL-1999; 99WO-US015570.
PR	10-JUL-1998; 98US-0092416P.
FR	12-MAY-1999; 99US-013870P.
PA	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PI	Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
DR	WPI; 2000-160827/14.
DR	P-PsDB; AAY77142.
PT	Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT	toxin serotypes A-G, is used for inducing an immune response against
PT	botulinum.
XX	

PS Example 3; Page 52; 54pp; English.  
XX  
CC The invention relates to novel vaccines that induce a protective immune  
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
CC DNA construct comprising a vector, and at least one nucleic acid fragment  
CC comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-  
CC G. In preferred embodiments of the invention, the vector is a Venezuelan  
CC equine encephalitis virus (VEE) replicon vector. Use of this vector  
CC results in the production of large amounts of a protein encoded by a  
CC sequence cloned into the replicon. The constructs are used to produce  
CC vaccines against botulism. The proteins can also be used as diagnostic  
CC tools for the diagnosis of botulism. The transformed host cells can be  
CC used to analyse the effectiveness of drugs and agents which inhibit toxin  
CC effects. The vaccine currently used against botulism is dangerous and  
CC expensive to produce, and contains formalin, which is very painful for  
CC the recipient. Also, the vaccine is incomplete, in that only 5 of the 7  
CC serotypes are represented in the formulation. The novel vaccine of  
CC overcomes these problems, as it is easily purified, and available in  
CC large quantities. It is also expressed in the lymph nodes for a better  
CC immune response. The present sequence represents DNA encoding native  
CC BoNTA heavy chain C-terminal fragment (Hc) used in an exemplification of  
CC the present invention  
XX  
SQ Sequence 1317 BP; 560 A; 108 C; 211 G; 438 T; 0 U; 0 Other;  
XX  
Query Match 2.1%; Score 42.4; DB 3; Length 1317;  
XX Best Local Similarity 49.5%; Pred. No. 2.2;  
XX Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
XX  
QY 1537 ATGTAAGAATTCATTAGATGCTGAAATAGACATGCTGCATTTCTGTCACCTT 1596  
DB 345 ATGGAAGATGATCTTATATATGTAATATGTAATATGTAATATGTAATATGTAAT 404  
QY 1597 ATCATAGAGAGTCATTAGCTAGAGAAACAAATCAATCTATGTAATTAGAGAACAA 1656  
DB 405 AAACAAAGAGTGTATTTTAAATACAGTCAAATGATTATATATGATTATATAACG 464  
QY 1657 GCTGTTTTCCTCAATATATAAATAAGAAAGAAACATGTAAGTCAAAATATTGT 1716  
DB 465 ATGGAATTTTGTACATCTCTATATATATGATTAAATTAATCTAAATTTATTAATGG 524  
QY 1717 TTAATCAGGTCATTGAGAAATCTATTAATAAGTATTGAAAT 1756  
DB 525 AAGATTAAATAGATCAAAACCAATTTCAAATTAGGTAAAT 564  
XX  
RESULT 42  
XX AAV30575  
XX AAV30575 standard; DNA; 1546 BP.  
XX  
AC AAV30575;  
XX  
DT 07-DEC-1998 (first entry)  
XX  
DE Clostridium botulinum toxin A fragment C gene in phisBcTA.  
XX  
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;  
XX botulism; ds.  
XX  
OS Clostridium botulinum; serotype A.  
XX Synthetic.  
XX  
FH Key Location/Qualifiers  
XX CDS 108..1496  
XX /\*tag= a  
XX  
XX MO9808540-A1.  
XX  
XX 05-MAR-1998.  
XX  
XX 28-AUG-1997; 97MO-US015394.  
XX

PR 28-AUG-1996; 96US-00704159.  
XX  
XX (OPH-) OPHIDIAN PHARM INC.  
XX  
XX Williams JA, Thalley BS;  
XX  
XX WPI; 1998-230234/20.  
XX  
XX P-PSDB; AAW68390.  
XX  
XX  
XX Host cell containing recombinant expression vector encoding Clostridium  
XX botulinum type B or E toxin - useful to treat humans and other animals at  
XX risk of intoxication with clostridial toxin.  
XX  
PS Example 28; Page 277-278; 428pp; English.  
XX  
XX This is the DNA sequence of the Clostridium botulinum serotype A toxin C  
XX fragment gene contained in plasmid phisBcTA. The encoded toxin A  
XX polypeptide (see AAW68390) has a histidine-tagged N-terminal extension.  
XX The vector was used to express native (i.e. non-fusion) soluble C  
XX fragment in Escherichia coli host cells. The invention relates to  
XX recombinant proteins derived from C. botulinum toxins. Methods are  
XX provided which allow for the isolation of soluble recombinant proteins  
XX free of significant endotoxin contamination. Preferred hosts for  
XX production of recombinant proteins are E. coli, insect cells and yeast  
XX cells. The recombinant toxins are used as immunogens for the production  
XX of vaccines and antitoxins that are useful in the treatment of humans and  
XX animals at risk of intoxication with clostridial toxin  
XX  
SQ Sequence 1546 BP; 629 A; 163 C; 256 G; 498 T; 0 U; 0 Other;  
XX  
Query Match 2.1%; Score 42.4; DB 2; Length 1546;  
XX Best Local Similarity 49.5%; Pred. No. 2.4; Indels 0; Gaps 0;  
XX Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
XX  
QY 1537 ATGTAAGAATTCATTAGATGCTGAAATAGACATGCTGCATTTCTGTCACCTT 1596  
DB 524 ATGGAAGATGATCTTATATATGTAATATGTAATATGTAATATGTAATATGTAAT 583  
QY 1597 ATCATAGAGAGTCATTAGCTAGAGAAACAAATCAATCTATGTAATTAGAGAACAA 1656  
DB 584 AAACAAAGAGTGTATTTTAAATACAGTCAAATGATTATATATGATTATATAACG 643  
QY 1657 GCTGTTTTCCTCAATATATAAATAAGAAAGAAACATGTAAGTCAAAATATTGT 1716  
DB 644 ATGGAATTTTGTACATCTCTATATATATGATTAAATTAATCTAAATTTATTAATGG 703  
QY 1717 TTAATCAGGTCATTGAGAAATCTATTAATAAGTATTGAAAT 1756  
DB 704 AAGATTAAATAGATCAAAACCAATTTCAAATTAGGTAAAT 743  
XX  
RESULT 43  
XX AA287219  
XX AA287219 standard; DNA; 1917 BP.  
XX  
AC AA287219;  
XX  
DT 08-MAY-2000 (first entry)  
XX  
DE DNA encoding native BoNT serotype A (BoNTA) N-terminal fragment (Hn).  
XX  
KW Botulinum neurotoxin; heavy chain; BoNT; serotype A; N-terminal fragment;  
XX Hn; Venezuelan equine encephalitis virus replicon; VEE; botulism;  
XX vaccine; diagnosis; drug screening; ds.  
XX  
OS Clostridium botulinum.  
XX  
FH Key Location/Qualifiers  
XX CDS 1..1221  
XX /\*tag= a  
XX /product= "Native botulinum neurotoxin serotype A (BoNTA)  
XX heavy chain N-terminal fragment"  
XX /note= "No stop codon given in the specification"  
XX

XX WO200002524-A2.  
 XX 20-JAN-2000.  
 XX 09-JUL-1999; 99MO-US015570.  
 XX 10-JUL-1998; 98US-0092416P.  
 XX 12-MAY-1999; 99US-0133870P.  
 XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L,  
 XX MPI; 2000-160827/14.  
 XX DR P-PSDB; AAY71141.  
 XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 XX toxin serotypes A-G, is used for inducing an immune response against  
 XX botulinum.  
 XX Example 3; Page 51; 54pp; English.  
 XX The invention relates to novel vaccines that induce a protective immune  
 XX response against Botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
 XX and G (BoNTA-BONTG). The vaccine of the invention is novel recombinant  
 XX DNA construct comprising a vector, and at least one nucleic acid fragment  
 XX comprising a C-terminal heavy chain fragment (Hc) from BoNT serotype A-  
 XX G. In preferred embodiments of the invention, the vector is a Venezuelan  
 XX equine encephalitis virus (VEE) replicon vector. Use of this vector  
 XX results in the production of large amounts of a protein encoded by a  
 XX sequence cloned into the replicon. The constructs are used to produce  
 XX vaccines against botulinism. The proteins can also be used as diagnostic  
 XX tools for the diagnosis of botulinism. The transformed host cells can be  
 XX used to analyse the effectiveness of drugs and agents which inhibit toxin  
 XX effects. The vaccine currently used against botulinism is dangerous and  
 XX expensive to produce, and contains formalin, which is very painful for  
 XX the recipient. Also, the vaccine is incomplete, in that only 5 of the 7  
 XX serotypes are represented in the formulation. The novel vaccine of  
 XX overcomes these problems, as it is easily purified, and available in  
 XX large quantities. It is also expressed in the lymph nodes for a better  
 XX immune response. The present sequence represents DNA encoding native  
 XX BoNTA heavy chain N-terminal fragment (Hn) used in an exemplification of  
 XX the present invention  
 XX  
 XX Sequence 1917 BP; 789 A; 185 C; 284 G; 659 T; 0 U; 0 Other;  
 XX  
 XX Query Match 2.1%; Score 42.4; DB 3; Length 1917;  
 XX Best Local Similarity 49.5%; Pred. No. 2.6;  
 XX Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
 QY 1537 ATGTAAGAATTCATTAGATTGGCTGAATAGACTGTCGATTTCTGCTGCT 1596  
 DB 1560 ATGGAAGAATTCATTAGATTGGCTGAATAGACTGTCGATTTCTGCTGCT 1619  
 QY 1597 ATCATAGAAGTCTAGTCTAGTGAAGCAAACTACATCTATGTATTAGAGAACAA 1656  
 DB 1620 AAAACAAGAGTGTATTAAATACAGCTCAAGATTAATATTCAGATTATTAACAG 1679  
 QY 1657 GCTGGTTTCTCAATTAATAAATAGAAAAAGAACCATGTAAGTCAAAATTTTGT 1716  
 DB 1680 ATGGAATTTTGTACTACTACTAATATATAGATTAAATTAATTTTATTAATGG 1739  
 QY 1717 TTAATCAGTCTAGTGAATCTATTAAGAATTTTGAAT 1756  
 DB 1740 AAGATTAAATGATCAAAACCAATTTCAAAATTTAGTTAAT 1779

RESULT 44  
 ID ADA71938 standard; DNA; 2000 BP.  
 XX  
 AC ADA71938;

XX 20-NOV-2003 (first entry)  
 XX Rice gene, SEQ ID 5263.  
 XX Plant; bacterial infection; fungal infection; viral infection; rice;  
 XX gene; db.  
 XX Oryza sativa.  
 XX WO2003000898-A1.  
 XX 03-JAN-2003.  
 XX 22-JUN-2001; 2001WO-IB001105.  
 XX 22-JUN-2001; 2001WO-IB001105.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,  
 XX Katagiri F, Quan S, Tao Y, Whitlam S, Xie Z, Zhu T, Zou G;  
 XX MPI; 2003-175290/17.  
 XX Identifying at least one gene involved in plant resistance or response to  
 XX pathogenic infection for conferring resistance or tolerance to a plant to  
 XX bacterial, fungal or viral infection by determining or detecting plant  
 XX gene expression.  
 XX Claim 27; SEQ ID NO 5263; 899bp; English.  
 XX The present invention relates to a method (M1) for identifying genes  
 XX involved in plant resistance or response to pathogenic infection. M1  
 XX comprises identifying a gene whose expression is significantly altered in  
 XX the incompatible interaction of plant gene expression relative to  
 XX expression of the gene in an uninfected plant, in a mutant plant that  
 XX does not express a gene associated with response to pathogenic infection,  
 XX or in a corresponding incompatible or compatible interaction. (M1) is  
 XX useful for conferring resistance to resistance or tolerance to a plant to  
 XX bacterial, fungal or viral infection. The present sequence was used to  
 XX illustrate the invention.  
 XX  
 XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
 XX  
 XX Query Match 2.1%; Score 42.4; DB 8; Length 2000;  
 XX Best Local Similarity 8.6%; Pred. No. 2.7;  
 XX Matches 61; Conservative 328; Mismatches 319; Indels 4; Gaps 2;  
 QY 1266 AGCTGGGCTCTTCAAGTAATCTTGGTTTTCACAAGCCCTAATTTTACTGCAATAT 1325  
 DB 918 WRTWSCWYTMWMMGAMRAYAYAMRRRRWTYKMSMRMYMTMTKAWMTWMTCMCAKMYMAT 859  
 QY 1326 AITTTGAATTCAGTATTAATTTCTCAATTTCCCATAGTCAATGATCAGACAAATACCT 1385  
 DB 858 GNAATWMMYVYTYCYATCTACKCKYKAMTMTWTTACAMATSTMRMAMGMRKRYK 799  
 QY 1386 CTCATGCAACACTGGCTTGTGTAATATCATATCATATATGAGA-GCTGTGCTTTAAGC 1444  
 DB 798 MKRATWMMWRCWKAGAMAMWMSRRIRMKKATIRYMGAMMTWMSRRKKSSTMRMG 739  
 QY 1445 GTAATGTTTTATATGACATTAAGCTTGGCTTCAATTAAGAAGGATTTAGAGCAATG 1504  
 DB 738 MGRMRSARVYOSRMKCACTKYASSABWTKRAKRSYRBRMYRMYRKRKGYRYRYSR 679  
 QY 1505 TGATACAGAAGCTTTTCTCCACAGGCTTCATATGTAAGAATTCATTAGTTGGCTGA 1564  
 DB 678 MTRARMSKRRKWAAGASMSKSCMMYRGARSMYSKSKCKCKKRYMTSTSTMGYGM 619  
 QY 1565 ATAGACTGATCGTCCATTTCTGCTGCTCATATATCAATAGAGTCAATGAGCAAGC 1624  
 DB 618 YSSYSMSWTSKMSYKMTCTWYTSMSKSTRSKMGSMGSMRYMRMKRKRKYMR 559

DNA construct comprising a vector and at least one nucleic acid fragment (HC) from BoNT serotypes A-G. In preferred embodiments of the invention, the vector is a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this vector results in the production of large amounts of a protein encoded by a sequence cloned into the replicon. The constructs are used to produce vaccines against botulism. The proteins can also be used as diagnostic tools for the diagnosis of botulism. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. The present sequence represents DNA encoding native BoNTA heavy chain used in an exemplification of the present invention

XX Sequence 2532 BP; 1038 A; 241 C; 404 G; 849 T; 0 U; 0 Other;

SO Query Match 2.1%; Score 42.4; DB 3; Length 2532;  
Best Local Similarity 49.5%; Pred. No. 3;  
Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 1537 ATGTAAGAATTCATTAGATTGGCTGAATAAGACTGATCTGTCCATTCTCTGCTCACTT 1596  
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 1560 ATGAAAGATGCACCTTAATTATGTGTAATAAATCTGGACCTTTACAGGATCACAAGAAAT 1619  
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| |||  
OY 1597 ATCATAGAAGATCATTAGCTAAGAACAAAACTACAATCTATGTATAATTAGAAGAACAA 1656  
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 1620 AAAACAAAGATGATGTTTTTAATAACAGCTCAAATGANTTAATATATACGATTATATAACAG 1679  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
OY 1657 GCTGTTTGCTCAATATAAAAATAAGAAAAAAAACCAATGTGAAGTCANAATATTGGT 1716  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 1680 ATGGAATTTTGTAACTATCACTAATATATAGATTAAATTAOTCTAAATATTATATAATGG 1739  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
OY 1717 TTATCAAGTCATTGAGATCTATTAAAAAGIATTTGAAT 1756  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 1740 AAGATTATATAGATCAAAAACCATTCCAATTAGGTAAT 1779  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: January 22, 2006, 04:09:00  
Job time : 813.081 secs



The page number (101)



c	37.6	1.9	83851	3	US-09-949-016-13847	Sequence 13847, A
99	37.6	1.9	144596	3	US-09-949-016-11749	Sequence 11749, A
100	37.6	1.9	144596	3	US-09-949-016-13035	Sequence 13035, A

## ALIGNMENTS

## RESULT 1

US-09-078-294-3

; Sequence 3, Application US/09078294

Patent No. 6265211

; GENERAL INFORMATION:

APPLICANT: Choo, Kong-Hong Andy

APPLICANT: Du Sart, Desiree

APPLICANT: Cancellia, Michael R.

TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE

FILE REFERENCE: Davies Col

CURRENT APPLICATION NUMBER: US/09/078.294

CURRENT FILING DATE: 1998-05-13

CURRENT FILING DATE: 12/15/2011  
NUMBER OF SEO ID NOS: 2

NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver 2.0

· SOFTWARE:  
· SEO IN NO 3

```

; SEQ ID NO 3
: LENGTH: 80595

```

LENGTH: 8  
TYPE: DNA

ORGANISM: Nucleotide sequence of HC-confirmed

; ORGANISM: N  
ITS - 09 - 078 - 394 - 3

Query match	100.0%;	Score 2001;	DB 3;	Length 80595;
-------------	---------	-------------	-------	---------------

QY	1	ATCTTTATTTGATAGAGAAAAAGAAAAAGAGTGAATAAATATATTTAACTGCAATTAAT	60
Db	78595	ATCTTTATTTGATAGAGAAAAAGAAAAAGAGTGAATAAATATATTTAACTGCAATTAAT	78654
QY	61	TCAGGACCATGATGTGCAGATGCAGAGAACTCAATTCAAACCAACGTAAAGTCAAAGGAA	120
Db	78655	TCAGGACCATGATGTGCAGATGCAGAGAACTCAATTCAAACCAACGTAAAGTCAAAGGAA	78714
QY	121	AATATATTTGGCTCATGTAACTTTCTCACAGAGGGCGAGATGGAAAGGGCTTTGGGAAC	180
Db	78715	AATATATTTGGCTCATGTAACTTTCTCACAGAGGGCGAGATGGAAAGGGCTTTGGGAAC	78774
QY	181	AAGAAATATGTTCTCAATTCCTAGAGAAATCAGAGATTAGTCCAGGATGGGTCACTTCCT	240
Db	78775	AAGAAATATGTTCTCAATTCCTAGAGAAATCAGAGATTAGTCCAGGATGGGTCACTTCCT	78834
QY	241	GTCCCTGAGTGTGTGTAGCGATGATGATCTTATGGAGAGAAAGATGCATGTTAGGA	300
Db	78835	GTCCCTGAGTGTGTGTGTAGCGATGATGATCTTATGGAGAGAAAGATGCATGTTAGGA	78894
QY	301	TGAAGATAGGGCTTAAGCAACAGAGGCGACATATTCATGCTAAAAATGTTTTT	360
Db	78895	TGAAGATAGGGCTTAAGCAACAGAGGCGACATATTCATGCTAAAAATGTTTTT	78954
QY	361	TTTGTATGCTCTCTTAATTTCACAAAATGCTTCCAAACAAAGTAGACACAGAGAAAAAGAC	420
Db	78955	TTTGTATGCTCTCTTAATTTCACAAAATGCTTCCAAACAAAGTAGACACAGAGAAAAAGAC	79014
QY	421	ATAGGACCTACTAGTGGGGTCTTTTATCTTAAAGCCTTGACTTGCTTTTCAACGTTA	480
Db	79015	ATAGGACCTACTAGTGGGGTCTTTTATCTTAAAGCCTTGACTTGCTTTTCAACGTTA	79074
QY	481	CTCACTGTGTGTAAGTGAAGGCATATATGCTGTAAAAAGCTTTCAGAGGTTTCTACTAAG	540
Db	79075	CTCACTGTGTGTGTAAGTGAAGGCATATATGCTGTAAAAAGCTTTCAGAGGTTTCTACTAAG	79134
QY	541	CTGGGTTCTTAATATAGGCTCTCTCCCAATTTCTGTGCTCACTGATGAATCTTTCTCT	600
Db	79135	CTGGGTTCTTAATATAGGCTCTCTCCCAATTTCTGTGCTCACTGATGAATCTTTCTCT	79194
QY	601	TTTCTCACTCTGAGGACGTAGTGGCTGTTTGTATGAACTGACCTTAGCTTCTTGGGTTT	660

Db	79195	TTCTCCACCTCTGGGACGTGGTGGCTGTTTGTAATGACCTCCCTAGCTTTGCTTGAGGTTT	79254
QY	661	TTTCTCTGGGGACAAATGCTCTTCAATATCTTAGACCAATAATAACTACAGCCACTGGGCA	720
Db	79255	TTTCTCTGGGGACAAATGCTCTTCAATATCTTAGACCAATAATAACTACAGCCACTGGGCA	79314
QY	721	GGCTCTCTCTCTCCAACTGGACCAATGTTCCAGGGCTCTTCACTTAAGTTAGTCAAG	780
Db	79315	GGCTCTCTCTCTCCAACTGGACCAATGTTCCAGGGCTCTTCACTTAAGTTAGTCAAG	79374
QY	781	CATTCTTGGCAAAAGAGGCCATGTTAAACAATAGCATTTAGAGCAATTGATCTTTTGG	840
Db	79375	CATTCTTGGCAAAAGAGGCCATGTTAAACAATAGCATTTAGAGCAATTGATCTTTTGG	79434
QY	841	ACATGTGTGAAGATCTATTCACATTTTGTAAATTAAGCAATCCCTATGGAACCAACAC	900
Db	79435	ACATGTGTGAAGATCTATTCACATTTTGTAAATTAAGCAATCCCTATGGAACCAACAC	79494
QY	901	GAACCTAAGCTGCTCTGGAATGACAGGGTGGCTCTCAATPAAGATGTTCTTAGAGCT	960
Db	79495	GAACCTAAGCTGCTCTGGAATGACAGGGTGGCTCTCAATPAAGATGTTCTTAGAGAGCT	79554
QY	961	GTAATTTTGGGCACTTAATCTATTCCTCCACTTAAGGGCACAGCACTGAATAACACAC	1020
Db	79555	GTAATTTTGGGCACTTAATCTATTCCTCCACTTAAGGGCACAGCACTGAATAACACAC	79614
QY	1021	TAACTTGTGATGCTCATGTAGTTAGTCTCAGGCACTGACGCTCAGAGTGAACCTGAC	1080
Db	79615	TAACTTGTGATGCTCATGTAGTTAGTCTCAGGCACTGACGCTCAGAGTGAACCTGAC	79674
QY	1081	CTCTTATGTGTGTCAGACCTTTTCTTCCCTCAGAAGTCACTGTGTTTCTGTGACTCTC	1140
Db	79675	CTCTTATGTGTGTCAGACCTTTTCTTCCCTCAGAAGTCACTGTGTTTCTGTGACTCTC	79734
QY	1141	CATAGGAACATCAGTCTCTGAATCTCAGACCAACATCTGAGATGAATGCTCTGACAC	1200
Db	79735	CATAGGAACATCAGTCTCTGAATCTCAGACCAACATCTGAGATGAATGCTCTGACAC	79794
QY	1201	GTCTCAGAAAGTGTCTCACCGCTGAAATCTCCAAGCGTGTGACACACCTGAGAGAAAT	1260
Db	79795	GTCTCAGAAAGTGTCTCACCGCTGAAATCTCCAAGCGTGTGACACACCTGAGAGAAAT	79854
QY	1261	GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTCAAGCCCCCTTAATTTTACTGCA	1320
Db	79855	GAGAAAGCTGGGCTCTTCAGGTAAATCTTGCTTTTCAAGCCCCCTTAATTTTACTGCA	79914
QY	1321	TAAATTTTGAATTCCTGATTAATTTTCTCAATTTTCCATPAAGCATCTACACACAT	1380
Db	79915	TAAATTTTGAATTCCTGATTAATTTTCTCAATTTTCCATPAAGCATCTACACACAT	79974
QY	1381	ACCCTTCATGCAACCTTGGCTTTGGCTAAATCATATCTAATATATGAGACTGCTCTT	1440
Db	79975	ACCCTTCATGCAACCTTGGCTTTGGCTAAATCATATCTAATATATGAGACTGCTCTT	80034
QY	1441	AAGGCTAAATGTTTATATGCACTAAGGCTCTTGCGCTTACATATAAAGGGGATATTGAC	1500
Db	80035	AAGGCTAAATGTTTATATGCACTAAGGCTCTTGCGCTTACATATAAAGGGGATATTGAC	80094
QY	1501	AATGTGATACGAAGTCTTTTCTCACAAGTCTCATATGTAAAGATTCATTTAGATTGGC	1560
Db	80095	AATGTGATACGAAGTCTTTTCTCACAAGTCTCATATGTAAAGATTCATTTAGATTGGC	80154
QY	1561	TGAATATGACTGATCTGTCACATTTCTGTGCACTTATCATTAAGAAAGTCATTAGCTAAG	1620
Db	80155	TGAATATGACTGATCTGTCACATTTCTGTGCACTTATCATTAAGAAAGTCATTAGCTAAG	80214
QY	1621	GAACAAAAACATCAATCTATGTATAATGAGAAACAAGCTGGTTTGTCTCAATATAAAAT	1680
Db	80215	GAACAAAAACATCAATCTATGTATAATGAGAAACAAGCTGGTTTGTCTCAATATAAAAT	80274
QY	1681	AAGAAAAAGAAACAATGTGAAGTCAAAATATTTGTTAATCAGGTCAATTGGAATCTAT	1740

Db 80275 AAGAAAAAACCATGTGAAGTCAAAATATTTTAAATCAGTCAATGAGAAATCTAT 80334  
Qy 1741 TAAAAATATTTGAATCTTTATGATGAGAACTATCTTGAAGTCAAGTGAAGTGGTGA 1800  
Db 80335 TAAAAATATTTGAATCTTTATGATGAGAACTATCTTGAAGTCAAGTGAAGTGGTGA 80394  
Qy 1801 CTTTGGGCTGTGGTCCCTACGATGAGAAAGAGGCTTTCATTAAGTCTTATATGTAC 1860  
Db 80395 CTTTGGGCTGTGGTCCCTACGATGAGAAAGAGGCTTTCATTAAGTCTTATATGTAC 80454  
Qy 1861 AGGTGCCAAGTTAAGTCCCAAGCTTGTCTTAAAGCATACGTGATTTTGTTAGACT 1920  
Db 80455 AGGTGCCAAGTTAAGTCCCAAGCTTGTCTTAAAGCATACGTGATTTTGTTAGACT 80514  
Qy 1921 TTTAGTGAAGTGAAGGAAATAAACAATCCCTCTGGGAAACTTCTCTCATCTTGGT 1980  
Db 80515 TTTAGTGAAGTGAAGGAAATAAACAATCCCTCTGGGAAACTTCTCTCATCTTGGT 80574  
Qy 1981 GAAGTCATCTGCAGAAATTC 2001  
Db 80575 GAAGTCATCTGCAGAAATTC 80595

## RESULT 2

US-09-078-294-6  
Sequence 6, Application US/09078294  
Patent No. 6265211  
GENERAL INFORMATION:  
APPLICANT: Choo, Kong-Hong Andy  
APPLICANT: Du Sart, Desiree  
APPLICANT: Cancilla, Michael R.  
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
FILE REFERENCE: Davies Col  
CURRENT APPLICATION NUMBER: US/09/078,294  
CURRENT FILING DATE: 1998-05-13  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 18443  
TYPE: DNA  
ORGANISM: HAC-F2 contig 2  
US-09-078-294-6

Query Match 99.2%; Score 1984.8; DB 3; Length 18443;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1997; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 ATCTTATTTGTATGAAAAAGAAAAAGAGTGAATAATATATTAAAGTGCATATAT 60  
Db 4292 ATCTTATTTGTATGAAAAAGAAAAAGAGTGAATAATATATTAAAGTGCATATAT 4351  
Qy 61 TCAGA-CCATGATTCAGATGACAGAACTCAATTCACCAACGTAAATCAAAAG 119  
Db 4352 TCAGAACCCTTGATTCAGATGACAGAACTCAATTCACCAACGTAAATCAAAAG 4411  
Qy 120 AAATATATTTGGCTCATGTAACCTTCTCAGAGAGGCGAGATGAAAGGGCTTTGGAA 179  
Db 4412 AAATATATTTGGCTCATGTAACCTTCTCAGAGAGGCGAGATGAAAGGGCTTTGGAA 4471  
Qy 180 CAAGAGAAATTTCTCAAAATCTTCAAGAAATCTAGGATTAATCCAGATGGGTCACTTCC 239  
Db 4472 CAAGAGAAATTTCTCAAAATCTTCAAGAAATCTAGGATTAATCCAGATGGGTCACTTCC 4531  
Qy 240 TGTCCCTGAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 299  
Db 4532 TGTCCCTGAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4591  
Qy 300 ATGAAGTAGGGCTAAACCAACAGGCGAGGCGCATATATCATGCTAAATATGTTT 359  
Db 4592 ATGAAGTAGGGCTAAACCAACAGGCGAGGCGCATATATCATGCTAAATATGTTT 4651  
Qy 360 TTTTATGTCCTCTTCAATTTCAAAATGCTTCCAAAGTAGCACAGAGAAAAAGAA 419  
Db TTTTATGTCCTCTTCAATTTCAAAATGCTTCCAAAGTAGCACAGAGAAAAAGAA 419

Db 4652 TTTTATGTCCTCTTCAATTTCAAAATGCTTCCAAAGTAGCACAGAGAAAAAGAA 4711  
Qy 420 CATAGGAACTTACTGATGGGCTTTTATTTTAAAGCTTTGATCTTTTCAAGCTT 479  
Db 4712 CATAGGAACTTACTGATGGGCTTTTATTTTAAAGCTTTGATCTTTTCAAGCTT 4771  
Qy 480 ACTCATGCTTTGATCTGAGGCGATATGAGGCTTGAAGAAAGCTTCTGAGGGTTCTACTAA 539  
Db 4772 ACTCATGCTTTGATCTGAGGCGATATGAGGCTTGAAGAAAGCTTCTGAGGGTTCTACTAA 4831  
Qy 540 GCTGGGCTCTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 599  
Db 4832 GCTGGGCTCTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4891  
Qy 600 TTTTCTCAGCTCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 659  
Db 4892 TTTTCTCAGCTCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 4951  
Qy 660 TTTTCTGAGGAAATGCTCTGAGATTAATCTAGAACCAATTAACCTAGAGGCACTGGGC 719  
Db 4952 TTTTCTGAGGAAATGCTCTGAGATTAATCTAGAACCAATTAACCTAGAGGCACTGGGC 5011  
Qy 720 AGGCTCTCTCTCTCCAACTGACCAATGTTCCAGGGCTTTTCACTTATGATGATCA 779  
Db 5012 AGGCTCTCTCTCTCCAACTGACCAATGTTCCAGGGCTTTTCACTTATGATGATCA 5071  
Qy 780 GCATTTCTTGGCAAAAGAAAGGCTTATTAACATTAACATTTAGCAATTTGATTTT 839  
Db 5072 GCATTTCTTGGCAAAAGAAAGGCTTATTAACATTAACATTTAGCAATTTGATTTT 5131  
Qy 840 GACATGTTAAAGATTAATCAATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAGAT 899  
Db 5132 GACATGTTAAAGATTAATCAATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAGAT 5191  
Qy 900 CGAATCAAGCTGCTCTGAGATGAGAGGCTCTCTCAATACAGATTTCTAGAGAGC 959  
Db 5192 CGAATCAAGCTGCTCTGAGATGAGAGGCTCTCTCAATACAGATTTCTAGAGAGC 5251  
Qy 960 TGTATTTTGGGCACTTAATTTCTCACTACTTAAAGGCAAGCACTGAATTAACACCA 1019  
Db 5252 TGTATTTTGGGCACTTAATTTCTCACTACTTAAAGGCAAGCACTGAATTAACACCA 5311  
Qy 1020 CTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1079  
Db 5312 CTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5371  
Qy 1080 CCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1139  
Db 5372 CCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5431  
Qy 1140 CCATAGGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1199  
Db 5432 CCATAGGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5491  
Qy 1200 AGTCCATGAAGTGTCTACCGCTGATCTCCAAAGGCTGACACCGGTGAGAGAGAA 1259  
Db 5492 AGTCCATGAAGTGTCTACCGCTGATCTCCAAAGGCTGACACCGGTGAGAGAGAA 5551  
Qy 1260 TGAGAAAGCTGGGCTTTTCAAGGTAATCTTCTTTTCAAGGCCCTTAATTTTACTGC 1319  
Db 5552 TGAGAAAGCTGGGCTTTTCAAGGTAATCTTCTTTTCAAGGCCCTTAATTTTACTGC 5611  
Qy 1320 ATATATATTTTGAATCACTGATTAATTTTCAAAATTTTCCATTAATGATCAACACAA 1379  
Db 5612 ATATATATTTTGAATCACTGATTAATTTTCAAAATTTTCCATTAATGATCAACACAA 5671  
Qy 1380 TACCTCTCATGAAACATTTGGCTTTGATATATATATATATATATATATATATATATAT 1439  
Db 5672 TACCTCTCATGAAACATTTGGCTTTGATATATATATATATATATATATATATATATAT 5731  
Qy 1440 TAAAGCTAAATGTTTATATATGATTAAGGCTCTTGGCTTAACATTAATAAGGGATTTGAG 1499  
Db 5732 TAAAGCTAAATGTTTATATATGATTAAGGCTCTTGGCTTAACATTAATAAGGGATTTGAG 5791



QY	1260	TGAGAAAGCTGGGCTCTTCAAGGTAAATCTTGCTTTTTCACAAAGCCCTTAATTTACTGC	1319
Db	79555	TGAGAAAGCTGGGCTCTTCAAGGTAAATCTTGCTTTTTCACAAAGCCCTTAATTTACTGC	79654
QY	1320	ATAATTAATTTTGAATTAATCACTGATTAATTTCTCAATATTTCCATTAAGTCATCTACACAA	1379
Db	79655	ATAATTAATTTTGAATTAATCACTGATTAATTTCTCAATATTTCCATTAAGTCATCTACACAA	79714
QY	1380	TACCTCTCATGCAACACTTGCGCTTTCCTAATACATATCTAATTATGAGAGCTGTGCTTCT	1439
Db	79715	TACCTCTCATGCAACACTTGCGCTTTCCTAATACATATCTAATTATGAGAGCTGTGCTTCT	79774
QY	1440	TAAACGTTAAAGTTTATATATGCACTAAGAGCTCTTGCTTACATATTAAGAGGATATTGAG	1499
Db	79775	TAAACGTTAAAGTTTATATATGCACTAAGAGCTCTTGCTTACATATTAAGAGGATATTGAG	79834
QY	1500	CAATGTGTAACAGAAAGCTTTTCTCCACAGGCTCATATGTAAAGAAATTCATTAGATTGG	1559
Db	79835	CAATGTGTAACAGAAAGCTTTTCTCCACAGGCTCATATGTAAAGAAATTCATTAGATTGG	79894
QY	1560	CTGAATATGACTGATCTGTCCATTTCTGTGCTCACTTATCATTAAGAAAGTCATTAGCTAA	1619
Db	79895	CTGAATATGACTGATCTGTCCATTTCTGTGCTCACTTATCATTAAGAAAGTCATTAGCTAA	79954
QY	1620	GGAACAAAACCTACATCTATGTATATTAGAAAGAACAGCTGGTTTGTCTCATATATAAAA	1679
Db	79955	GGAACAAAACCTACATCTATGTATATTAGAAAGAACAGCTGGTTTGTCTCATATATAAAA	80014
QY	1680	TAAAGAAAAGAACCATATGTAAAGTCAAAATATTGTTTATATGAGGTCATTGGAATCTAA	1739
Db	80015	TAAAGAAAAGAACCATATGTAAAGTCAAAATATTGTTTATATGAGGTCATTGGAATCTAA	80074
QY	1740	TTAAAGAGTATTTGAATTTCTTTATGATGAGAACATATTTGACTCAAGTGGACAGTGGTGA	1799
Db	80075	TTAAAGAGTATTTGAATTTCTTTATGATGAGAACATATTTGACTCAAGTGGACAGTGGTGA	80134
QY	1800	GCTTTTGGCGCTGTGTCCTCTACGTAGAAAGAGAGGCTTTGTCTATAAAGCTTATATATGTA	1859
Db	80135	GCTTTTGGCGCTGTGTCCTCTACGTAGAAAGAGAGGCTTTGTCTATAAAGCTTATATATGTA	80194
QY	1860	CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTTTAAAGCATCTGATTTTGG	1911
Db	80195	CAGGTGCCAAGTTAAGTGCCCAAGCTTGCTTTAAAGCATCTGATTTTGG	80246
RESULT 4			
US-09-078-294-22			
; Sequence 22, Application US/09078294			
; Patent No. 6265211			
; GENERAL INFORMATION:			
; APPLICANT: Choo, Kong-Hong Andy			
; APPLICANT: Du Sart, Destiree			
; APPLICANT: Cancelli, Michael R.			
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE			
; FILE REFERENCE: Davies Col			
; CURRENT APPLICATION NUMBER: US/09/078,294			
; CURRENT FILING DATE: 1998-05-13			
; NUMBER OF SEQ ID NOS: 29			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 22			
; LENGTH: 300			
; TYPE: DNA			
; ORGANISM: BAC-F2 contig 47 fragment 2			
US-09-078-294-22			

		Query Match	4.9%	Score 97.6	DB 3	Length 300
		Best Local Similarity	66.7%	Pred. No. 1,1e-17		
		Matches 168	Conservative 18	Mismatches 62	Indels 4	Gaps 4
OY	1753	GAATTCCTTATCATGAAACTATCTTGACCTCAAGTGCACAGTGTGACCTTTGGCCTG				18112
		:::::	:	:	:	:
Db	1	GAATTCCTCGWATTAKAACATATCTTGCCTCAAATTSCTTGGTAGGTAACCTGGCCTG				60

OY	1813	TGGTCCCTACG-TGAAGAAGAGCTTTGTCAATAAG-TCCTATXNTGTACAGGCGCCAAG	1870
Dd	61	TGCTCCTCTTGACTTAATGAGAGCTTTGTCAATATGATCATMTGTGTATCTKGTGCCTAG	120
OY	1871	TT-AAGTGCCCMAGCTTGCTCTTAAAGCATCTGAATTTTGTATTAGACTTTTAAGTGAA	1929
Dd	121	TTTGAATGCCCCCTGCCTTCTGTSTTCTWGGGCTTACTKGATTTWGGGGTATNACAATCAATKAAA	180
OY	1930	CTGAAGGGAATTAACAATCCCTCTGTGGAGAACCTTCTCTCCATCTTTGTGGAAGTCAAT	1989
Dd	181	YTSAAGAAGTC-TTCTCCTCCCGYGAGAAATTTCTCTCCTCCCTCGGAGAACTCTTT	239
OY	1990	CTGCCAGAAATTC	2001
Dd	240	CTSCCGAATTC	251

```

1      RESULT 5
2      US-08-232-463-14
3      ; Sequence 14, Application US/08232463
4      ; Patent No. 5670367
5      ; GENERAL INFORMATION:
6      ; APPLICANT: DORNER, F.
7      ; APPLICANT: SCHEIFLINGER, F.
8      ; APPLICANT: FALKNER, F. G.
9      ; TITLE OF INVENTION: RECOMBINANT FOXP10X VIRUS
10     ; NUMBER OF SEQUENCES: 52
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Foley & Lardner
13     ; STREET: 1800 Diagonal Road, Suite 500
14     ; CITY: Alexandria
15     ; STATE: VA
16     ; COUNTRY: USA
17     ; ZIP: 22313-0299
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: Patentln Release #1.0, Version #1.25
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/08/232,463
25     ; FILING DATE:
26     ; CLASSIFICATION: 435
27     ; PRIOR APPLICATION DATA:
28     ; APPLICATION NUMBER: US/07/935,313
29     ; FILING DATE:
30     ; APPLICATION NUMBER: EP 91 114 300.6
31     ; FILING DATE: 26-AUG-1991
32     ; ATTORNEY/AGENT INFORMATION:
33     ; NAME: BENT, Stephen A.
34     ; REGISTRATION NUMBER: 29,768
35     ; REFERENCE/DOCKET NUMBER: 30472/114 IMPTU
36     ; TELECOMMUNICATION INFORMATION:
37     ; TELEPHONE: (703) 836-9300
38     ; TELEFAX: (703) 683-4109
39     ; TELEX: 899149
40     ; INFORMATION FOR SEQ ID NO: 14:
41     ; SEQUENCE CHARACTERISTICS:
42     ; LENGTH: 7218 base pairs
43     ; TYPE: nucleic acid
44     ; STRANDEDNESS: single
45     ; TOPOLOGY: linear
46     ; IMMEDIATE SOURCE:
47     ; CLONE: pTZgpt-F18
48     ; US-08-232-463-14

```

[illegible]

```

QY 468 TTTTCACAGTCTTCTGCTTGTACCTGAGCCATATGCTGTAAGCTTCTGAG 527
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1106 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1165
QY 528 GGTTCCTACTAAGCTGGTCTTATATAGCTCTCCATTTCTGTGCTCCTAG 587
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1166 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1225
QY 588 TGAATCTTCTCTTCTCTACCTCTGGAGCTGTGTGTATGAGACTGCTAGCT 647
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1226 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1285
QY 648 TTGCTTTGGTTTCTTCTGAGCAATGCTTCAATATCTAGCAATATACTAG 707
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1345
QY 708 AGCCACTGGCGAGCTCTCTCTCCAACTGAGACATGTTCCAGGCTCTTACCTT 767
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1346 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1405
QY 768 AGTTAGCTCAACATCTT 787
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1406 YYYYYYYYYYYYYYYYYY 1425

```

```

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F15

```

```

US-08-232-463-14
Query Match      2.9%; Score 58.4; DB 2; Length 7218;
Best Local Similarity 1.4%; Pred. No. 9.5e-06;
Matches 5; Conservative 216; Mismatches 127; Indels 0; Gaps 0;
QY 8 TTTGATGAGAAAAAGAGAGTGAATAATATATTAAGTCATATAGTTCAGAC 67
    |||||
Db 1441 TTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382
QY 68 CATGATTCAGAGTACAGAACTCAATTCAACCAAGTAAAGTCAAAAGAAATAT 127
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322
QY 128 TGGCTCATGTAACCTTCTCAGAGAGGAGAGTGAAGAGGAGCTTTGGAACAAGAA 187
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262
QY 188 TTGTTCTCAAAATTCAGAAATCTAGATTAAGTCAAGATGAGTCACTTCTGCTG 247
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202
QY 248 AGGTGCTGATGAGTGTAGTCTTATGAGAGAGAGAGTGCATGTTAGATGAAGT 307
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142
QY 308 AGGCGTAAGCAAAAGGAGGAGGCACTATATCATGCTAATAATGG 355
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1094

```

```

RESULT 7
US-09-513-999C-19284
; Sequence 19284, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 19284
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 425
; OTHER INFORMATION: r=a or g
US-09-513-999C-19284

```

```

Query Match      2.5%; Score 49.4; DB 3; Length 431;
Best Local Similarity 71.4%; Pred. No. 0.00081;
Matches 65; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 70 TGGATGCAAGTACGAGAAATCAATTCAAACCAAGTAAAGTCAAAAGAAATATATG 129
    |||||
Db 231 TTGGTTGTAAGTACGAGAAACCAACTCAAAATTAAGCTTAGGCCAAAAGAGAAATATAG 290
QY 130 GCTCATGTAACCTTCTCAGAGAGAGGAGGAG 160
    |||||
Db 291 GTTCATGAAACCAAACTAAGAAAGGAGAG 321

```

```

RESULT 8
US-09-806-708B-22/c

```



```
/ Sequence 22, Application US/09806708B
/ Patent No. 6784342
/ GENERAL INFORMATION:
/ APPLICANT: The University of British Columbia
/ TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
/ FILE REFERENCE: 4810-58741
/ CURRENT APPLICATION NUMBER: US/09/806,708B
/ PRIOR FILING DATE: 2001-04-03
/ PRIOR APPLICATION NUMBER: US 60/147,133
/ PRIOR FILING DATE: 1999-08-04
/ NUMBER OF SEQ ID NOS: 23
/ SOFTWARE: Patent version 3.0
/ SEQ ID NO 22
/ LENGTH: 1141
/ TYPE: DNA
/ ORGANISM: Artificial sequence
/ FEATURE:
/ NAME/KEY: promoter
/ LOCATION: (1)..(1141)
/ OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match
Best Local Similarity 12.8%; Pred. No. 0.0066;
Matches 75; Conservative 205; Mismatches 303; Indels 3; Gaps 2;

QY 807 TAAACAATGACATTTAGCAATGATTTCTTTGACATGTTGAGATCTATTCACATTT 866
DB 926 KANNNNNNNNYATYACVRAATNNKMAHMKWTHGHSKERTHHRTGRTKYNNNNNN 867
QY 867 TGTAAATTAAGCATTCCTCCATGGAACCAACGAGCAAGTGGCTCCCTGGAATGAC 926
DB 866 AATVYVYHHAARMMAMWMTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 807
QY 927 GTGGCTCCCTCATATACAGATGTTCTAGAGAGCTGATTTTGGGCACTTAACTATTTCTC 986
DB 806 NNNTWCHYTTANABCBYRANNNNAARMAATCANNYHAATTTHTDWCYKTMNTWYND 747
QY 987 ACTACTTAAAGGACACAGCATGAATTAACACCACTAAGTTTGCATGCTCCAGTAACT 1046
DB 746 MTMTBTTTTRNMTSTNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 689
QY 1047 TCTCAGGAGTGGAGCTGAGAGTGAAGTGAAGTCTTATGCTGCTCAGCCTTTCTTC 1106
DB 688 RRTNNNTTMMRRNMTTKTRWYSTTRHHYTGATNNNNNNNNNNNNNNNNNNNN 629
QY 1107 CTTCAGAGTCAAGCTGTGTTTCTGCTGCTCTCATAGACATCATGCTCGAATCTTC 1166
DB 628 MKWTMGDGTIVAKYKMRDITCTIYDVADSVWVYAMNMRGRDYTRNNNTYCKSYAH 569
QY 1167 AGACCAACATCTGAGTGAAGTGTCTCTGACAGTCTTGAAGTGTCTGAGCTGGAT 1226
DB 568 SVYWMNNMMWYRYSARNNSSMARWTTTRNNNNNNNNNNNNNNNNNNNNNNNN 509
QY 1227 CTCCAAAGCTGTGACACACCGTGAAGAGAAATGAGAAAGCTGGCTCTTCAGTAAAT 1286
DB 508 YMMWKMAWBT-TTVYDSMCNNAASMRGNMRKMAANNDAAGDHATTYMGANN 450
QY 1287 CTGCTTTTTCACACGCCCTTATTTTCTGCATATATTTTGAATTCATCATATAT 1346
DB 449 MRRAMMMAMMAWRAVCCNNNNNNRACVHKKHGMWTKYMKAKACNNNNBEAMTRVA 390
QY 1347 TCTACAATTTTCCATAGCATCTACACACAACTAACCTCTCATGTC 1392
DB 389 WMYSHDTTNTDMMWMTSDMBWMTYTVDYTMBAWNNNNNNNNNNNNNNNNNN 344

RESULT 9
US-09-949-016-203912
/ Sequence 203912, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
```

```
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 203912
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-203912
```

```
Query Match
Best Local Similarity 49.7%; Pred. No. 0.0054;
Matches 172; Conservative 0; Mismatches 172; Indels 2; Gaps 2;

QY 1446 TAAATGTTTATGACATTAAGCTCTTGGCTTACATATATAAAGGGTATGACAATGT 1505
DB 116 TTATGCTTTGTTTCTTAAACTGTATGTTTAAACAAAATATGATTTGAAAAAAT 175
QY 1506 GATPACAGAGCTTTCTTCACAGGCTCATATGTAAGAAATTCATTAATGCTGGA 1565
DB 176 GAACCAACAGGTGACCAAAAAGTACATCTGTTACAGACATCA-CTGCTTACA 234
QY 1566 TAGCTGATCTGCTCATTTCTGCTCACTTATCATAGGAAGTCAATTAGCTAAGA 1625
DB 235 ACAATTCACATATAGATTTAAGATTTAACTGATCTTAAGTATCAGAACATTA 294
QY 1626 AAAATCAGATCTATGATTAAGAAACAGAGCTGTTTGGCTCATATTAATAAATA 1685
DB 295 AAAAAGCTCAAGTACCTATTTAAATTAATCGTGCTTCAATATATATTTCTTAA 354
QY 1686 AAGAAGAAACATGTGA-AGTCAAAATATTTTATATCAGTCAATGAGATCTATTA 1744
DB 355 AATGAAAAACAGTAAAGACACACATTTTATTACTGCAACACATGATTAATATGA 414
QY 1745 AAGTATTTGAATCTTTATGATGAGAACTATCTTGAATCTGAATGGA 1790
DB 415 AACTTTTGAATTTTCTTTTCTTTAAGACATTTTCTCTAGAGTA 460

RESULT 10
US-09-949-016-5761/c
/ Sequence 5761, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 5761
/ LENGTH: 3338
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-5761
```

Query Match 2.3%; Score 46.8; DB 3; Length 3338;  
Best Local Similarity 49.7%; Pred. No. 0.013;  
Matches 172; Conservative 0; Mismatches 172; Indels 2; Gaps 2;

QY 1446 TAAATGTTTATATGACACTAAGGCTTGGCTTACATATATAAAGGGGTATGAGCATGT 1505  
DB 3142 TTAATGCTTTGTTTCTAAACCTTGATGTTTTTAAACAAAATGATTTGAAAAGAT 3083

QY 1506 GATACAGAGTCTTTCTGCACAGTCTCATATGTAAGAAATCTAGTAGTGGCTGAAA 1565  
DB 3082 GAACCAACAAGCTGACCAACAAAAGTCAATCTTTGTTACAGACATCA-CTGCTCTTACA 3024

QY 1566 TAGACTGATCTGTCATTTCTGCTCACTTATCATAGAAATCATTAAGTCAAGACA 1625  
DB 3023 ACAATTCACATAATAGATTAAAGATTAACTTCATCTCAATGATATGACAACTTAAAA 2964

QY 1626 AAAACTACATCTATGTTATTAAGAACAGAGTGGTTTGCATATATAAATAGAA 1685  
DB 2963 AAAAAAAGTCAATGACCTATTTAAATTAATGCTTCAATTAATATTTCTTTAAA 2904

QY 1686 AAGAAACCATGTGAA-AGTCAAAATATTTGTTAATCAAGTCAATGAGATCTATTAA 1744  
DB 2903 AATGAAAAACGTAAGAGACACATTTTATTACTACACACTGAATTAATATGTA 2844

QY 1745 AAGTATTTGAATTTCTTATGATGAGAACTATCTTGACTCAAGTGA 1790  
DB 2843 AACTTTTGAATTTTTTTTTTTTCTTAGACATTTTCTCTAGAGTA 2798

RESULT 11  
US-09-949-016-17503/c  
; Sequence 17503; Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17503  
; LENGTH: 38119  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17503

Query Match 2.3%; Score 46.8; DB 3; Length 38119;  
Best Local Similarity 49.7%; Pred. No. 0.047;  
Matches 172; Conservative 0; Mismatches 172; Indels 2; Gaps 2;

QY 1446 TAAATGTTTATATGACACTAAGGCTTGGCTTACATATATAAAGGGGTATGAGCATGT 1505  
DB 35922 TTAATGCTTTGTTTCTAAACCTTGATGTTTTTAAACAAAATGATTTGAAAAGAT 35863

QY 1506 GATACAGAGTCTTTCTGCACAGTCTCATATGTAAGAAATCTAGTAGTGGCTGAAA 1565  
DB 35862 GAACCAACAAGCTGACCAACAAAAGTCAATCTTTGTTACAGACATCA-CTGCTCTTACA 35804

QY 1566 TAGACTGATCTGTCATTTCTGCTCACTTATCATAGAAATCTTAAGTCAAGACA 1625  
DB 35803 ACAATTCACATAATAGATTAAAGATTAACTTCATCTCAATGATATGACAACTTAAAA 35744

QY 1626 AAAACTACATCTATGTTATTAAGAACAGAGTGGTTTGCATATATAAATAGAA 1685  
DB 35743 AAAAAAAGTCAATGACCTATTTAAATTAATGCTTCCATTAATATATTTCTTTAAA 35684

QY 1686 AAAGAAACCATGTGAA-AGTCAAAATATTTGTTAATCAGGTCATTGAGAAATCTATTA 1744  
DB 35683 AATGAAAAACGTAAGAGACACATTTTATTACTCACAACTGATATATATGTA 35624

QY 1745 AAGTATTTGAATTTCTTATGATGAGAACTATCTTCACTCAAGTGA 1790  
DB 35623 AACTTTTGAATTTTTTTCTTTTACGACATTTTCTCTAGAGTA 35578

RESULT 12  
US-09-806-708B-22  
; Sequence 22; Application US/09806708B  
; Patent No. 6784342  
; GENERAL INFORMATION:  
; APPLICANT: The University of British Columbia  
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
; FILE REFERENCE: 4810-58741  
; CURRENT APPLICATION NUMBER: US/09/806,708B  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: US 60/147,133  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 1141  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: (1)..(1141)  
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters  
US-09-806-708B-22

Query Match 2.3%; Score 46; DB 3; Length 1141;  
Best Local Similarity 13.1%; Pred. No. 0.013;  
Matches 77; Conservative 223; Mismatches 278; Indels 11; Gaps 2;

QY 1174 CATCTGAGTAGTAGTCTCTCTGACAGTCTAGAGTTGTTCTACCGGTGATCTCAAA 1233  
DB 195 MYRYTDRMSBKRNMYGMBWKMSYDTYYTWWDDMCCKXVRRWTRGRMNNYMA 254

QY 1234 GCGTGTACACACCGTGAAGAAATGAGAAAGCTGGGCTTCAAGTAAATCTTGCT 1293  
DB 255 MBTARRRRYNNGMTBAMAYRWTNNNNNNNAKMKRAKAYGMBABVNSTCTWKSKT 314

QY 1294 TTTTCAAGCCCCCTTAATTTTCTGCTATATTTGATTAATTCATCTGATTAATTTTACAA 1353  
DB 315 KVRTSCWANNCPAGDANKDKWKSAAAGVYNNNNNNNNNNWYKARHBAWDVWHSW 374

QY 1354 TTTTCCCATPAGTCATCT-----ACACACAATACCTCTCATGCAACACTTGCT 1403  
DB 375 KKHANAAHYSKKRTBYRKRTVNNNNNGTTWKKMMAMVWMDMDWBGTYNNNNNGRT 434

QY 1404 TTGCTAATACATATCTATTTATGAGAGCTGTGCTTTTAAAGCTTAATGTTTATATGAC 1463  
DB 435 YGWTNKKKMMTYKWKANNCKRAMDHKTCTHNNNTWKKTYNNNNCYMKSMGTGKSHR 494

QY 1464 TAAAGCTCTGGCTTACATATAAAGGGGTATGAGCAATGATACAGAAATCTTTCT 1523  
DB 495 BAAAVT-WYWWWRRYAAANNNDYWKACITWKKYBVCWMMNYAAWYKSSMNTYS 553

QY 1524 CCACAGGCTCATATGTAAGAAATCATTAAGTGGCTGAATATGACTGATCTGCATT 1583  
DB 554 RYRMRKTNNSSWRSDTRSMGRANNYABABHGYKMTNRWMSHTWBHBRGAHMYMB 613

QY 1584 TCTGCTCACTTATCATTAAGAAAGTCACTTAAGAAACAAAACATCAATCTATGTA 1643  
DB 614 MWYBAKCHCKMAYRAKKYAGAGSNNNNNNNNNNNNNNNNNNATCARDDVYASRWYMAN 673

QY 1644 ATTTAAGAAACAGCTGTTTGGTCTCATATATAAATAAGAAAAACCATGTGAAG 1703  
DB 674 AKWYKYBAANNAYYTHANNMWGCWNNATDTRTYWKNNNNNNNAGTWNKNNNNNAKNSA 733





```

1 APPLICANT: WILLIAMS, JAMES A.
2 TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
3 TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
4 NUMBER OF SEQUENCES: 30
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
7 STREET: 220 MONTGOMERY STREET, SUITE 2200
8 CITY: SAN FRANCISCO
9 STATE: CALIFORNIA
10 COUNTRY: UNITED STATES OF AMERICA
11 ZIP: 94104
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentin Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/084,517
20 FILING DATE:
21
22 CLASSIFICATION:
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 08/
25 FILING DATE: 16-MAR-1995
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 08/329,154
28 FILING DATE: 25-OCT-1994
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: US 08/161,907
31 FILING DATE: 02-DEC-1993
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 07/985,321
34 FILING DATE: 04-DEC-1992
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: US 07/429,791
37 FILING DATE: 31-OCT-1989
38 ATTORNEY/AGENT INFORMATION:
39 NAME: CARROLL, PETER G.
40 REGISTRATION NUMBER: 32,837
41 REFERENCE/DOCKET NUMBER: OPD-01610
42 TELECOMMUNICATION INFORMATION:
43 TELEPHONE: (415) 705-8410
44 TELEFAX: (415) 397-8338
45 INFORMATION FOR SEQ ID NO: 27:
46 SEQUENCE CHARACTERISTICS:
47 LENGTH: 3891 base pairs
48 TYPE: nucleic acid
49 STRANDEDNESS: double
50 TOPOLOGY: linear
51 MOLECULE TYPE: DNA (genomic)
52 FEATURE:
53 NAME/KEY: CDS
54 LOCATION: 1..3888
55 US-09-084-517-27
56
57 Query Match 2.1%; Score 42.4; DB 3; Length 3891;
58 Best Local Similarity 49.5%; Pred. No. 0.26;
59 Matches 109; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
60
61 QY 1537 ATGTAAAGAAATTCATTAGATTTGGCTGAATAGACTGATCTGTCATTCTCGTCACTT 15366
62 Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
63 Db 2919 ATGGAAGATTCACCTTAATTATGTTAGTGTAATTAATCTGCACTTTACAGATACTCAGAAAT 29787
64
65 QY 1597 ATCTAAGAGAGTCATTAGCTAGTAAGGACAAAACTACAACTATGATTAATTAGAAACAA 16566
66 Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 Db 2979 AAAACCAAGGTAGTCTTTTAAATATACGTCAAAATGATTAATATACAGATTATATTAACAG 30389
68
69 QY 1657 GCTGGTTTTCCTCAATATTAATAATTAAGAAAAAGAAACCATGTGAAGTCAAAATATTTGT 17166
70 Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 Db 3039 ATGGAATTTTGTACATCTACTAATATATATGATTAAATTAATCTCTAAATTTATATTAATGG 30988
72
73 QY 1717 TTATCTAGGTCATTGAGAAATCTATTAATAAAGTATTTGAAT 1756
74 Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
75 Db 3099 AAGATTATATAGATCAAAAACCAATTTCAAAATTTTAGGTAAT 3138

```



```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203451
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-203451

```

```

Query Match      2.1%; Score 42; DB 3; Length 601;
Best Local Similarity 48.3%; Pred. No. 0.13;
Matches 114; Conservative 1; Mismatches 121; Indels 0; Gaps 0;

```

```

QY 672 CAATGCTTCAATTTCTTCAAGCAATAAATACTAGACGACGAGGCTCTTCTTC 731
    |||||
DB 269 CAATCTTTCAAGCTTAAAAAAMAAAAAAMAAAAAAGGGCTCTTAATGATCTTC 328
    |||||
QY 732 CTCACACTGGACCATGTTCCAGGGCTCTTCACTTGAAGTCAAGATCTTGGCA 791
    |||||
DB 329 TCCCTACTGCTGGCTTCCCTGCGCTCAATTTCTCTATTTCTCCCTACACACCTTTGAC 388
    |||||
QY 792 AAGAAAGGCTTAGTTAACAATAAGACATTTAGCAATTTGTTTGAATGTTGTA 851
    |||||
DB 389 CCAGCCACACTGAATGCTTGACGCTCCCAACAGCTAATCTGTTAATGCTTTATG 448
    |||||
QY 852 GATCTATTCACATTTTGTATTAAGATTCCTCCCTATGAGAAACACACAGACTAA 907
    |||||
DB 449 GTTTCAAACACATCAGGCTAGAAATGCTTCTCCCTTTCTAATAACATCTACTCA 504
    |||||

```

## RESULT 23

```

; Sequence 203452, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203452
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-203452

```

```

Query Match      2.1%; Score 42; DB 3; Length 601;
Best Local Similarity 48.3%; Pred. No. 0.13;
Matches 114; Conservative 1; Mismatches 121; Indels 0; Gaps 0;

```

```

QY 672 CAATGCTTCAATTTCTTCAAGCAATAAATACTAGACGACGAGGCTCTTCTTC 731
    |||||

```

```

DB 277 CAATCTTCAAGCTTAAAAAAMAAAAAAMAAAAAAGGGCTCTGAATGATCTTC 336
    |||||
QY 732 CTCACACTGGACCATGTTCCAGGGCTCTTCACTTGAAGTCAAGATCTTGGCA 791
    |||||
DB 337 TCCCTACTGCTGGCTTCCCTGCGCTCAATTTCTCTATTTCTCCCTACACACCTTTGAC 396
    |||||
QY 792 AAGAAAGGCTTAGTTAACAATAAGACATTTCTGCAATTTGTTTGAATGTTGTA 851
    |||||
DB 397 CCAGCCACACTGAATGCTTGACGCTCCCAACAGCTAATCTGTTAATGCTTTATG 456
    |||||
QY 852 GATCTATTCACATTTTGTATTAAGATTCCTCCCTATGAGAAACACACAGACTAA 907
    |||||
DB 457 GTTTCAAACACATCAGGCTAGAAATGCTTCTCCCTTTCTAATAACATCTACTCA 512
    |||||

```

## RESULT 24

```

US-09-949-016-14773
; Sequence 14773, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14773
; LENGTH: 113060
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14773

```

```

Query Match      2.1%; Score 42; DB 3; Length 113060;
Best Local Similarity 56.5%; Pred. No. 2;
Matches 78; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

```

QY 8 TTTGTATGCAAAAAGAAAAGAGTGAATAATATTTATTTAGCTGATTTAGTCAAGAC 67
    |||||
DB 111243 TTTATTTAGCTCAATTAATAAGAAATTTATTTATCACTGATTTAGCAGGATGC 111302
    |||||
QY 68 CATGATTCGACGTGCAGAAATCAATCAACCAAGTAAGTCAAAAGGAAATATAT 127
    |||||
DB 111303 TATCTTTTCAAGTGCAGAAAATTTCAATTAACGGGCTTAAACGTTAAAGTGAAATTTG 111362
    |||||
QY 128 TGGCTCATGTAACTTCT 145
    |||||
DB 111363 AGGCTCAAGTAAGTACT 111380
    |||||

```

## RESULT 25

```

US-09-949-016-14774
; Sequence 14774, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498

```

```

Query Match      2.1%; Score 42; DB 3; Length 113060;
Best Local Similarity 56.5%; Pred. No. 2;
Matches 78; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```





OY 1446 TAAATGTTTATATGCACTAAGCGCTTGCGTTACATATAAAAAGGGGTATTGAGCAATGT 1505  
| | | | | | | | | | | | | | | | |  
Db 3118 TTAAATGCTTGTGTTCTTAAACTGTATGTTTTTTAAACAATAATGATTTGAAGAAGAT 3059

Query Match 2.1%; Score 41.2; DB 6; Length 3312;  
 Best Local Similarity 48.8%; Pred. No. 0.53;  
 Matches 169; Conservative 0; Mismatches 173; Indels 4; Gaps 2;



```
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14043
LENGTH: 312470
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(312470)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14043
```

```
Query Match          2.0%; Score 41; DB 3; Length 312470;
Best Local Similarity 47.8%; Pred. No. 6.5; Indels 130; Gaps 0;
Matches 119; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
```

```
QY 1536 TATGTAAGAAATTCATTAGATTGGCTGAATAGACTGATCTGCAATTCCTGCTCACT 1595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19287 TATCAAAATTCATAAACAGTCACCTTAATAAGTAAGCAATTTTCTGTAATTN 19228

QY 1596 TATCATAAAGAGACTCTTACCTAGAGACAAACAACTCAATCTATGTATTAAGAAAC 1655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19227 AATTTTAAAAATGAAATTAAGAGTCACAAAATTAATAAATGTTAACTAACAAAAAC 19168

QY 1656 AGCTGGTTTGTCTCAATATATAAATAAGAAACCAATGGAAGTCAAAATATTG 1715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19167 AATTATATCTTGACATTAATAAATGACCAACCACTTATTAAGACTAGATGATG 19108

QY 1716 TTTAATCAGGTCAATGAGATCTATTAATAAAGTATTGATCTTATGATGAACTAT 1775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19107 TTTCCACTATCTTCAGGAAATTTACTTTAAATATTTTATATACGATGACCAATCAT 19048

QY 1776 CTGACTCA 1784
    ||| ||| |||
DB 19047 GTTGTCTAA 19039
```

```
RESULT 33
US-09-949-016-17484/c
; Sequence 17484, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17484
; LENGTH: 92155
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17484
```

```
Query Match          2.0%; Score 40.8; DB 3; Length 92155;
Best Local Similarity 48.3%; Pred. No. 3.9; Indels 122; Gaps 0;
Matches 114; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
```

```
QY 672 CAATGCTTCAGATTATCTTACAGCAAAATTAACAGCACTGGCCAGGCTCTTC 731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23679 CAATCTTCAGCTTAAAAAAGGAGGCTCTGAAATGATCTTC 23620

QY 732 CTCCACTGGACCAATGTTCCAGGGCTCTTCACTTATGTTAGTCAGCATTTGGCA 791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23619 TCTCACTGGCTCGGCTCTCCCTGCTCAATTCCTTATCTCCACACACCTTTGAC 23560

QY 792 AAAGAAAGGCTAGTAAACAATAGACATCTGCAATGATCTTTTGACATGTTGTA 851
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23559 CCAAGCACTGAAATGCTTGGACCTCTCAACAGTAATCTGTTATAGCTTTATG 23500

QY 852 GATCTATTCACATTTTGTATTAAGACATTTCCCTATGGAACCAACAGAACTAA 907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23499 GTTTCAAACACATCAGGCTAGAAATGCTTCTCCCTTTCTTAATATACATCTACTCA 23444
```

```
RESULT 34
US-09-949-016-86304/c
; Sequence 86304, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86304
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-86304
```

```
Query Match          2.0%; Score 40.6; DB 3; Length 601;
Best Local Similarity 50.3%; Pred. No. 0.32; Indels 99; Gaps 0;
Matches 100; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
```

```
QY 1564 AATAGACTGATCTGCTCAATTTCTGCTCACTTATCATTAAGAGATCACTAGAGAA 1623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 515 AATTATATCATTTATTTATTTTGAACATATTCATAAATAATCTTTATTTTAA 456

QY 1624 CAAAACTACATCTATGTAATTAGAGACAAAGCTGTTTCTCAATATTAATAAG 1683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 455 TATAGAAAACATTTTGTATATTAATAATTAATAACATATCTTAATATTAATAAATA 396

QY 1684 AAAAAGAAACCAATGGAAGTCAAAATATTTTAAATGAGCAATGGAATCTATTA 1743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 395 TAAACATTTATATTAATATATATGAAATTAATATTTATTTATTAATATTAATTAAC 336

QY 1744 AAAGTATTTGAATCTTTA 1762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 335 AGTATATTTTATATATTTTA 317
```

```
RESULT 35
US-09-949-016-86305/c
; Sequence 86305, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```



```

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ. ID NOS: 28208
; SEQ ID NO 10353
; LENGTH: 252
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-10353

Query Match
Best Local Similarity 53.1%; Score 40.4; DB 3; Length 252;
Pred. No. 0.23; Mismatches 76; Indels 0; Gaps 0;
Matches 86; Conservative 0;

QY 1602 AAGGAAGTCATTGCTAGAGACAAACATCAATCTATGTAATTAGAGACAGCTGG 1661
DB 228 AATGTCATATGATGATATATTAATAGCAATTCACAAAAATAGGATGAAAAAGCTAACTGA 169
QY 1662 TTTTGCTCATATATTAATAAGAAAAAGAAACATGTGAAGTCAAAATATTTGTTTAT 1721
DB 168 TTTTATATGAAGATATACAGAGAGAAAGATCTTGAAAAATGAAAAATTAATAAAAT 109
QY 1722 CAGGTCATTGAGATCTATTAAAAAGTATTGAATCTTTAT 1763
DB 108 TAGATGTGTACAAAAATTTATGATGATGATGATTTT 67

RESULT 39
US-09-573-080A-118/c
; Sequence 118, Application US/09573080A
; GENERAL INFORMATION:
; APPLICANT: JOAN, KNOLL
; APPLICANT: KOGAN, PETER
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATI
; FILE REFERENCE: 30307
; CURRENT APPLICATION NUMBER: US/09/573,080A
; CURRENT FILING DATE: 2000-05-16
; NUMBER OF SEQ. ID NOS: 479
; SOFTWARE: Patent version 3.0
; SEQ ID NO 118
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: (1)..(476)
; PUBLICATION INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Jurka, J, Walichewicz, J, Milosavljevic, A
; TITLE: Protocypic sequences for human repetitive DNA
; JOURNAL: Journal of Molecular Evolution
; VOLUME: 35
; ISSUE: 4
; PAGES: 286-291
; DATE: 1992-10-
; DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
; DATABASE ENTRY DATE: 1996-01-26
; DATABASE ENTRY DATE: 1996-01-26
US-09-573-080A-118

Query Match
Best Local Similarity 2.0%; Score 39.6; DB 3; Length 476;
Pred. No. 0.54; Mismatches 1; Indels 10; Gaps 0;
Matches 45; Conservative 1;

QY 85 AGAACTCAATTCACCAAGTACGTAAGCAAGAAATATATTTGCTCATGTAC 140
DB 461 AGAACTCAATTCACCAAGTACGTAAGCAAGAAATATATTTGCTCATGTAC 406
```

```

RESULT 40
US-09-949-002-645
; Sequence 645, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J, Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: C1000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 645
; LENGTH: 56374
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-645

Query Match
Best Local Similarity 2.0%; Score 39.6; DB 3; Length 56374;
Pred. No. 6.7; Mismatches 129; Indels 0; Gaps 0;
Matches 117; Conservative 0;

QY 1496 TGAGCAGTGTGATACAGAGCTTTCTCCACAGTCTCATATGTAAAGATTCATTAGA 1555
DB 29627 TTTAAATGTTATTTCTCTAGATTATATACAGTTATATATACCAATTAAATTC 29686
QY 1556 TTGGCTGAATAGACTGATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1615
DB 29687 CAATATCTTAAACTTGTCAAAACATGCTTAAATTTCTGCAAAAATGCTGGAACG 29746
QY 1616 CTAGAGACAAACATCAATCTATGTAATTGAGAGACAGCTGTTTGCATATATA 1675
DB 29747 TTAGCCAGAAAAATTTGATGAGAAAAAGAGCATTCTTGTAAAGTATTAATA 29806
QY 1676 AAAATAGAAAAAGAAACCATGGAAGTCAAAATTTGTTAATCAGGTCATTGAGA 1735
DB 29807 TATTATTAAGATGAGATTTATGATGTCTCACAAAAATTTGCTAGGGGCTCCTAGCA 29866
QY 1736 TCTATT 1741
DB 29867 CTAAAT 29872

RESULT 41
US-09-949-002-774
; Sequence 774, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J, Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: C1000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 774
; LENGTH: 56375
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-774

Query Match
Best Local Similarity 2.0%; Score 39.6; DB 3; Length 56375;
Pred. No. 6.7; Mismatches 129; Indels 0; Gaps 0;
Matches 117; Conservative 0;
```





QY 1727 CATTGAGATCTATTAAAGTATT 1751  
DB 1269 AAGTACAAATATTATTAATAAANT 1293

## RESULT 44

US-09-350-756-8  
; Sequence 8, Application US/09350756  
; Patent No. 6495143  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP Riid 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; EARLIER FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 8  
; LENGTH: 1987  
; TYPE: DNA  
; ORGANISM: Clostridium botulinum  
; FEATURE:  
US-09-350-756-8

Query Match 2.0%; Score 39.2; DB 3; Length 1987;  
Best Local Similarity 51.1%; Pred. No. 1.5;  
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1589 GCTCAGTATCATTAAGAACTCTAGCTAGAGACAAAACCTACATCTATGTAATTAG 1648  
DB 825 GGTCTAATAATATATAGTAACAATGGTTAGCAAGTTAATACACAGATTGATCTAAT 884  
QY 1649 AAGAACAGCTGTTTGTCTCAATATTAATAAAGAAAGAACCATGTGAAGTCAAA 1708  
DB 885 AAGAAAAAATGAAGAACTTTAGAAATCAAGCAGACAAAGGCTATATATAA 944  
QY 1709 ATATTGTTTAATCAGTCAATGAGAACTTATTAATAAAGTATTGAATCTTTATGATGA 1768  
DB 945 CTATCAGTATATCAATATATCTAGAGAAAGAAAAATTAATTAATTTTAATATGATGA 1004

## RESULT 45

US-09-350-756-7  
; Sequence 7, Application US/09350756  
; Patent No. 6495143  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP Riid 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; EARLIER FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 7  
; LENGTH: 2452  
; TYPE: DNA

; ORGANISM: Clostridium botulinum  
; FEATURE:  
US-09-350-756-7

Query Match 2.0%; Score 39.2; DB 3; Length 2452;  
Best Local Similarity 51.1%; Pred. No. 1.7;  
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1589 GCTCAGTATCATTAAGAACTCTAGCTAGAGACAAAACCTACATCTATGTAATTAG 1648  
DB 705 GGTCTAATAATATATAGTAACAATGGTTAGCAAGTTAATACACAGATTGATCTAAT 764  
QY 1649 AAGAACAGCTGTTTGTCTCAATATTAATAAAGAAAGAACCATGTGAAGTCAAA 1708  
DB 765 AAGAAAAAATGAAGAACTTTAGAAATCAAGCAGACAAAGGCTATATATAA 824  
QY 1709 ATATTGTTTAATCAGTCAATGAGAACTTATTAATAAAGTATTGAATCTTTATGATGA 1768  
DB 825 CTATCAGTATATCAATATATCTAGAGAAAGAAAAATTAATTAATTTTAATATGATGA 884

Search completed: January 22, 2006, 21:30:21  
Job time : 255.225 secs

This Page Blank (uspto)